Conformational Analysis of Intrinsically Disordered Proteins Using Mass Spectrometry-Based Approaches

By

Mohammed Ali Hussain Al-Naqshabandi

B.Sc., Chemistry, 2004, Salahaddin University, Erbil, Iraq

M.Sc., Analytical Chemistry, 2009, Salahaddin University, Erbil, Iraq

Submitted to the graduate degree program in Chemistry and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Chairperson – David D. Weis, Ph.D.

_____

Cindy Berrie, Ph.D.

_____

Roberto De Guzman, Ph.D.

_____

Carey Johnson, Ph.D.

_____

Michael Wang, Ph.D.

Date Defended: July 18, 2016

The Dissertation Committee for Mohammed Ali Hussain Al-Naqshabandi

certifies that this is the approved version of the following dissertation:

Conformational Analysis of Intrinsically Disordered Proteins Using Mass Spectrometry-Based Approaches

_____

Chairperson – David D. Weis, Ph.D.

Date approved: July 20, 2016

ABSTRACT

Intrinsically disordered proteins (IDPs) have regions that are highly flexible and lack stable secondary or tertiary structure. Recently, there has been a growing interest in IDPs due to their important roles in many biological processes and functions. Although many studies have shown that IDPs participate in functional interactions, less is known about the structural details of the interactions. Over the past two decades, mass spectrometry (MS) has become a powerful technique for biophysical characterization of proteins due to its high sensitivity and variety of choices for sample preparation and instrumentation. Here I present the application of two important mass spectrometry-based approaches: hydrogen exchange (HX) and fast photochemical oxidation of proteins (FPOP) to study disordered proteins.

HX-MS was applied to better understand the mechanism of calcineurin activation. Calcineurin is a heterodimeric phosphatase that plays essential roles in cellular processes. Previous work has established that at high calcium concentration, calmodulin binds calcium ions, resulting in calmodulin binding to the intrinsically disordered regulatory domain of calcineurin. Calmodulin binding causes release of the autoinhibitory domain from the active site, activating calcineurin. My results with full-length calcineurin demonstrate that the regulatory domain is unstructured in the absence of calmodulin, while it folds upon binding to calmodulin. This result confirms previous work on the isolated regulatory-autoinhibitory domain construct. Additionally, I have observed calmodulin-induced changes in peptides located in other domains of calcineurin. Finally, and surprisingly, I found no changes in the structuring of the calcineurin autoinhibitory domain upon calmodulin binding. I present results from all regions of the calcineurin to describe the mechanism of calcineurin activation. I also propose a new model of calcineurin activation upon calmodulin binding.

The degree of structure measurement in IDPs can provide important information about the mechanisms by which IDPs undergo coupled folding and binding. Different approaches to quantify the degree of structure in IDPs using millisecond HX were explored. A quench-flow device, built in-house, for HX labeling on the millisecond timescale was employed. It is essentially impossible to determine the degree of structure without having an accurate unprotected reference state. The interaction domains of the activator for thyroid and retinoid receptors (ACTR) and the CREB binding protein (CBP) were used as model IDPs to explore the best approach to produce an unprotected reference state for millisecond HX. ACTR is a near-random coil IDP that has some residual helicity while CBP is a molten globular IDP that transiently becomes unstructured as revealed by NMR and HX-MS. The approaches explored to obtain an unprotected reference state in HX experiments were chemical exchange calculations, addition of denaturing agents, and millisecond HX labeling of peptic peptides obtained from the IDP. It was found that peptic reference peptides can be used as an accurate unprotected reference state for determining the degree of the structure.

Due to its fast labeling timeframe, FPOP might reveal states of IDPs that are undetectable by HX. The application of the FPOP technique for characterizing IDPs was also evaluated. To explore the applicability of this technique for studying IDPs, ACTR and CBP as model systems that co-fold upon binding were used. The FPOP technique was utilized to study ACTR and CBP in their free and bound forms. The data show that FPOP provided useful information to compare two states of IDPs. The usefulness and limitations of FPOP analysis to characterize and localize regions of protein-protein interactions involving IDPs are illustrated.

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor, Prof. David Weis, for his intellectual contributions to my development as a scientist. He has given me tremendous help and guidance in the past five years. He always had an open door to me for no matter what the reason was. I admire his extreme patience, enthusiasm for science, and organization skills. This Ph.D. dissertation could not have been accomplished without his mentoring and support. I will be forever grateful for helping me to learn and master the concepts presented here in this dissertation. I am also grateful and proud to be one of his graduate students.

I would like to thank the members of my dissertation defense committee: Prof. Carey Johnson, Prof. Cindy Berrie, Prof. Roberto De Guzman, and Prof. Michael Wang, for taking their time to evaluate my research. I want to thank Prof. Heather Desaire, Prof. Mario Rivera, and Prof. Robert Dunn, for their participation in my oral exam. I would also like to thank the entire faculty in the department of Chemistry for offering coursework that helped me throughout this research project. Thanks also go to my colleagues, Dr. Theodore Keppel, Farai Rusinga, Tyler Hagman, Dr. Ranajoy Majumdar and Jayant Arora for the fun and support in the lab.

I am grateful to the NIH/NIGMS Biomedical Mass Spectrometry Resource at the School of Medicine at Washington University in St. Louis for giving me the opportunity to perform part of this project in their labs. I would also like to thank Prof. Michael Gross, Dr. Hao Zhang, Dr. Henry Rohrs, and Ben Niu at the Washington University in St. Louis for the support during fast photochemical oxidation of proteins project. I would like to thank Prof. Peter Wright (Scripps Research Institute, La Jolla, CA) for providing the ACTR expression plasmid and Dr. Magnus Kjaergaard for providing the ACTR secondary chemical shift data in tabular form. I am also grateful to Prof. Treavor Creamer (University of Kentucky) for providing the proteins:

calcineurin, RD-AID-CT, and calmodulin for calcineurin project. I am also thankful to Erik Cook from Creamer lab for calcineurin activity assay data.

I thank Dr. Todd Williams for use of the KU Mass Spectrometry and Analytical Proteomics Laboratory. I would also like to thank Justin Douglas for assistance with ^{1}H NMR data collection of deuterated guanidine. I thank Agilent Technologies for an equipment loan. The financial support for this research that was from the National Science Foundation CAREER award to D. D. Weis (MCB-1149538) is gratefully acknowledged. I would also like to thank the Higher Committee for Education Development in Iraq for the scholarship given to me to pursue my Ph.D. study. I would like to thank Soran University for all the support during my graduate study.

Last but not least, I would like to thank my parents, Ali Hussain and Jameela Shamssadin, for their constant support, encouragement, and inspiring me to do the best I can. I am especially thankful to my supportive wife, Savana Jamal, for her love, encouragement, patience, and care. Without her support, this project could not have been successfully accomplished. I also thank my lovely children: Aland, Malak, and Rahand for filling my life with love and happiness. I also thank my brother, sisters, and family for believing in me to achieve my educational goals. I dedicate this dissertation to the memory of my grandfather, Hussain Abdulla and to the memory of my uncle, Hassan Abdulla for their support and inspiration.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter One

Introduction and overview

1.1 Intrinsically disordered proteins

1.1.1 Classification of disorder and the protein quartet

Proteins are a class of biological molecules that are involved in nearly all cell processes. Although some degree of flexibility was allowed in a protein molecule, it has been pointed out that protein function is directly related to protein's three dimensional structure.[1] In 1894, Emil Fischer proposed the lock-and-key model to describe the molecular mechanism of enzymatic activity.[2] This model has been more supported after successful resolution of many protein structures by X-ray crystallography, which results in reinforcing a static view of functional proteins. The structure-function paradigm became the center of many major directions of protein science like enzymatic catalysis, proteomics, protein misfolding, allosteric interactions, protein self-organization, bio-medicine, and protein engineering.[3] In fact, even the most well-defined proteins are considered dynamic systems having different degrees of conformational flexibility. These well-defined proteins undergo structural rearrangements of relatively small scale and relatively quickly.[4, 5] However, there is a different class of proteins that exhibit a high degree of conformational flexibility and lack a structured conformation under physiological conditions, but are still functional.[6-10] In the literature, several terms have been used to describe this class of proteins like partially folded,[11] natively disordered,[12] intrinsically unfolded,[13] natively denatured,[14] and intrinsically disordered.[15] In this dissertation, these proteins will be referred to as intrinsically disordered proteins (IDPs). Generally, IDPs lack bulky hydrophobic amino acids that make up the hydrophobic core in forming a well-defined structure.[16, 17] Hence, the IDPs functionality arises in a different manner as compared to the classical structure-function model of ordered proteins.[3]

In general, based on the degree of disorder, proteins can be categorized into a quartet of states.[18] These states are ordered, pre-molten globule, molten globule, and random coil. It has been proposed that protein function can arise from any of these four states and from transitions between them.[18] Figure 1.1 shows a schematic representation of the continuum model of protein structure. Folded proteins tend to have well-defined secondary and tertiary structural elements that have formed from the folding of the protein.[19] The stable folded state of a protein limits the number and variability of different conformers of folded proteins.[18, 19] The structure-disorder space of a protein is considered as a continuum, and there is no distinct boundary among disordered proteins on the continuum. The molten globule is a state that is located in the middle of the continuum. Molten globular proteins are loosely defined proteins that retain secondary structural elements but lack a stable tertiary fold.[20-22] The pre-molten globule is a state that has high conformational flexibility which makes the secondary structure of the protein undergo transient disruptions.[23, 24] The transient disruption of the secondary structures causes the protein hydrodynamic radius to become even larger, which in turn allows IDPs to bind to targets from a larger distance.[23] Although proteins can be theoretically categorized into random coils that have no secondary structure, proteins never become complete random coils and always have some residual structure.[25]

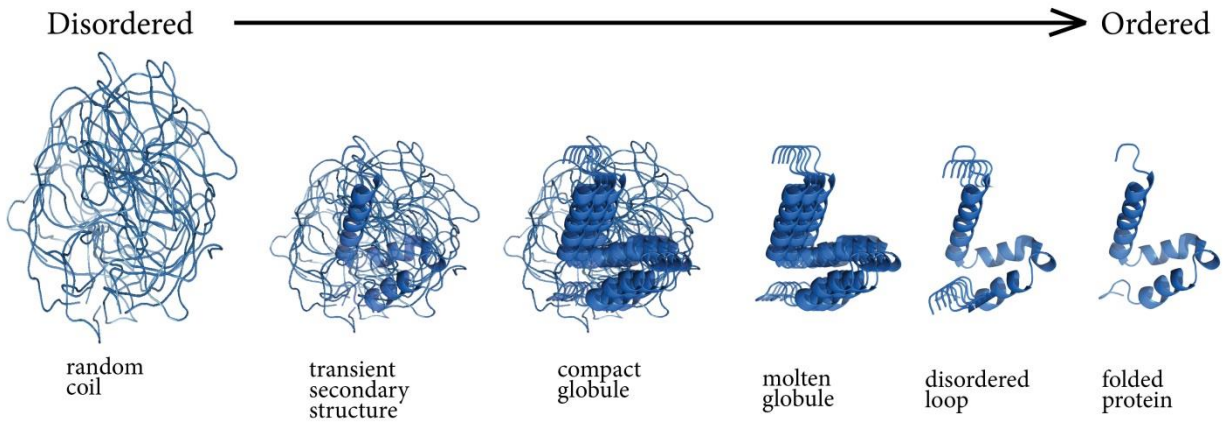


Figure 1.1. Schematic representation of the continuum model of protein structure. The continuum of conformational states ranging from highly dynamic and expanded conformational ensemble (left) to compact and fully folded globular states (right). Dynamically disordered states are represented by lines and folded structures are represented by cartoons. Redrawn from Ref. (18).

1.1.2 Abundance of IDPs in various proteomes

Several prediction algorithms have been applied to proteome databases to identify the prevalence of disordered proteins.[16, 26-29] These predictions revealed that IDPs and intrinsically disordered regions are prevalent in eukaryotic genomes.[26, 30-32] For example, a study revealed that a 44% of human protein-coding genes contain regions of predicted extended disorder of >30 amino acids in length.[33] The study also showed that 6% of all protein-coding genes of the human genome do not have any function annotation in their description, and these genes contain more predicted IDPs. Other studies showed that eukaryotic proteomes exhibit more disorder than prokaryotic or archaea proteomes with *C. elegans, A. thaliana, S. cerevisiae,* and *D. melanogaster* predicted to have 52-67% of proteins with intrinsically disordered regions, while bacteria and archaea were predicted to have 16-45 and 26-51% of proteins with disordered regions, respectively.[26, 30] These predictions of high frequency of IDPs in eukaryotic proteomes

may indicate the IDPs importance to perform necessary functions in multicellular eukaryotic organisms.[34]

1.1.3 IDPs and human diseases

IDPs have an important role in many biological processes and functions. The biological functions include transcriptional regulation, cellular signaling, post-translational modification, and cellular recognition.[10, 35, 36] Dysregulation of IDPs in many biological processes has been linked to a number of human diseases.[37-40] One of the disease states is cancer.[15] There is a strong association between IDPs and cancer-associated proteins, and examples of the IDPs associated with the cancer disease are p53,[41] Bcl-X_L and Bcl-2,[42] c-Fos,[43] and a thyroid cancer associated protein, TC-1.[44] Furthermore, there are several IDPs involved in neurodegenerative diseases like Alzheimer's disease,[45] Down's syndrome,[46] Parkinson's disease,[47] and prion diseases.[47] The IDPs involved in these disease states are tau-protein, α-synuclein, amyloid-β, and PeP^{SC}. The neurogenerative diseases are also called conformational diseases because the disease occurs due to conformational changes, misfolding, and aggregation of the proteins linked to the disease states.[3] In addition, there are IDPs, α-synuclein, for example, associated with neurodegenerative diseases because of dysregulation in protein functionality rather than only conformational disorders.[3] The funtions attributed to α-synuclein include binding fatty acids and metal ions, regulation of transporters, and regulation of specific enzymes.[47]

1.1.4 Coupled folding and binding mechanism

One of the properties of IDPs is their highly dynamic conformations that mediate protein-protein interaction through the phenomenon of coupled folding and binding.[21, 48, 49] In this type of interaction, folding of a disordered protein occurs upon engagement with its binding partner.

Therefore, there are two necessary events in coupled folding and binding mechanism: a conformational change and a binding step.[50] There has been considerable debate about the order of these events.[50] Two extreme mechanisms have been proposed. Conformational selection is one of the mechanisms of binding proposed in IDP interactions.[51, 52] According to this mode of binding, transiently folded states, which can exist within conformational ensembles of disordered proteins, can be selected by the binding partner to allow binding. The other mode of binding is the induced folding mode.[51, 52] In this case, the conformation flexibility of a disordered protein can ease the binding with its binding partner. This binding leads to folding the disordered protein. Although differentiation of these two mechanisms is an ongoing challenge, it is possible that processes might have elements of both mechanisms depending upon experimental conditions including protein concentration.[50] One example of processes having elements of both mechanisms is c-Myb and CBP. A recent paper showed that a disordered protein, c-Myb, binds to the binding site on the KIX domain of CBP by a mechanism that combines elements of both conformational selection and induced folding.[53] Away from which mechanism is taking place, another property of IDPs is their ability to return to the highly flexibile conformations after the completion of a specific function and their tendency to obtain different conformations depending on the environmental particularities.[3]

1.1.5 Model IDPs: ACTR and CBP

In this dissertation, two model IDPs are studied extensively in chapters 2, 3, and 4. These proteins are a binding domain of human p160 coactivator, activator for thyroid and retinoid receptors (residues 1023-1093, hereafter referred to as ACTR) and the nuclear coactivator binding domain (NCBD) of mouse cAMP-responsive element (CRE)-binding (CREB) binding protein (residues 2059-2117, hereafter referred to as CBP). There are extensive data

characterizing ACTR and CBP proteins in both isolated and complex forms.[21, 54-56] ACTR is a near-random coil IDP that transiently become structured (so-called residual structure),[54-56] while CBP is a molten globular IDP that transiently becomes unstructured[55-57] (see Figures 1.2a and b). As reported in the case of the interaction between some IDPs, the interaction between ACTR and CBP occurs through a process of mutual induced folding.[21] The isolated forms of ACTR and CBP are disordered, while they form a tight and structured complex when bound to each other (see Figure 1.2c) with a K_d value of 3.4×10^{-8} M.[21] In chapter 3, isolated forms of ACTR and CBP have been used as model systems to identify the best approach to measure residual structure of IDPs. In chapter 4, both isolated and complex forms of ACTR and CBP have been used to evaluate the application of fast photochemical oxidation of protein mass spectrometry to studying conformational dynamics of IDPs.

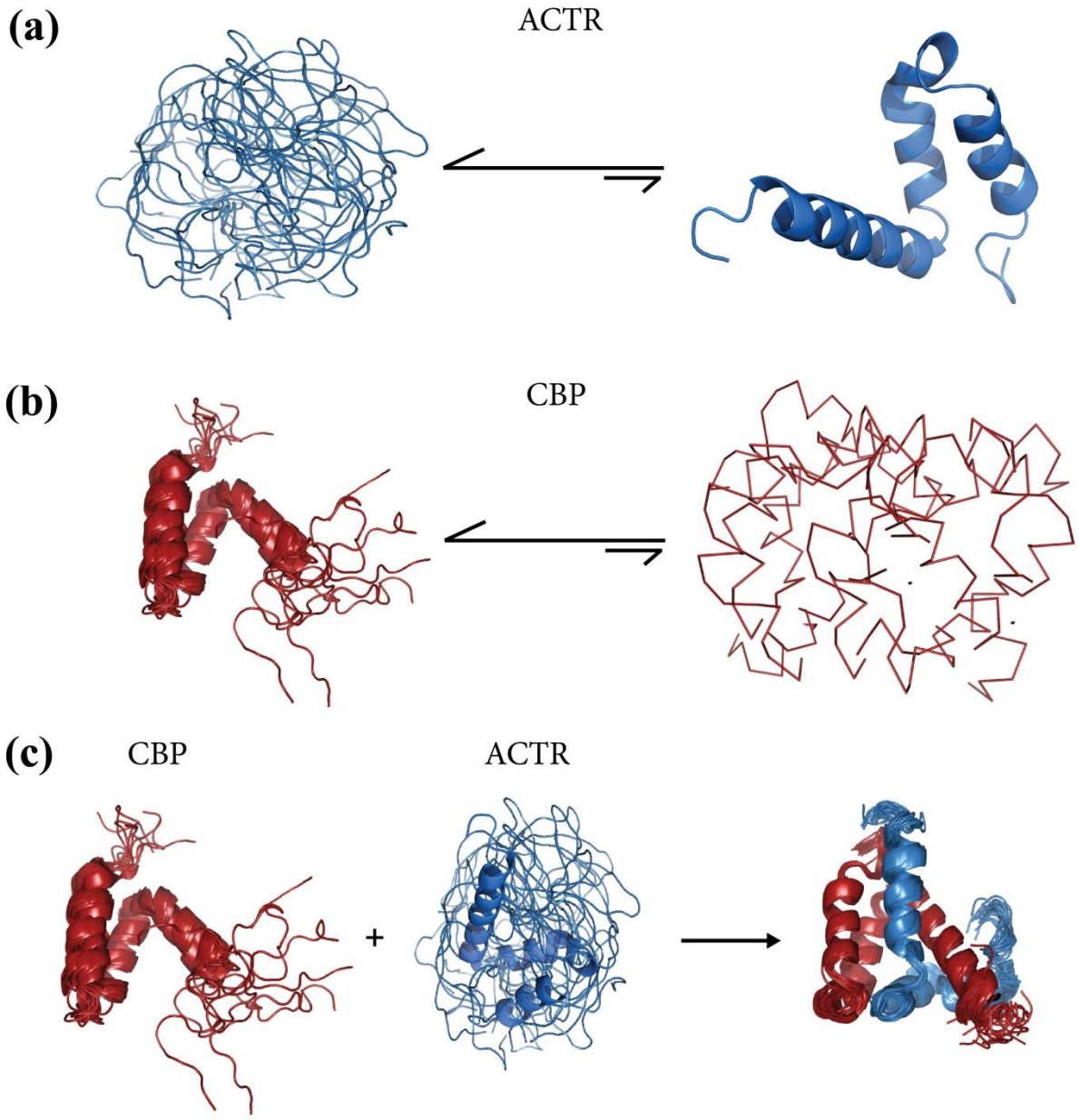


Figure 1.2. Cartoon representation of isolated (a) ACTR and (b) CBP and (c) their complex form. ACTR is a near-random coil IDP that transiently becomes structured. CBP is a molten globular IDP that transiently becomes unstructured. ACTR and CBP form a structured complex when they bind to each other. The depictions of ACTR and CBP are modified from PDB entry 1KBH (21).

1.1.6 Calcineurin

In chapter 2 of this dissertation, I studied human calcineurin (CaN), an IDP that has two unstructured regulatory regions. CaN is a heterodimeric serine/threonine phosphatase activated by calcium and calmodulin (CaM) binding.[58-61] CaN plays important roles in various cellular and biological processes such as T-cell activation, muscle contraction, neuronal signaling, and cardiac growth.[60] CaN has two subunits (A and B chains). There are four distinct functional domains in CaN A: a catalytic domain, a CaN B-binding domain, a calmodulin-binding domain, and an auto-inhibitory domain[62] . In its free form, CaN has two unstructured regulatory regions, the C-terminal domain and the regulatory domain (except for a short, structured autoinhibitory region, see section 2.1 for more details). Depending on the calcium concentrations in the cell, CaN undergoes sequential conformational changes to become fully activated.[63] Chapter 2 of this dissertation begins with a review of CaN activation models. Next, hydrogen exchange mass spectrometry has been used to understand the mechanism of CaN activation in solution after calmodulin binding.

1.2 Detecting and characterizing IDPs

Many techniques have been used to identify and characterize IDPs. Some of the techniques are considered indirect because they do not directly provide structural information and only provide global properties of a particular protein under study. These techniques include resistance to heat and chemical denaturation, limited proteolysis, and differential scanning calorimetry. Others techniques provide information on the size or hydrodynamic property of IDPs, such as gel filtration chromatography, dynamic light scattering, small-angle X-ray scattering, and analytical ultracentrifugation. Since the conformational behavior of IDPs is different from that of ordered proteins, many spectroscopic techniques have been used to provide

both steady-state and dynamic structural information on IDPs at low and high resolution. Some of the computational and experimental techniques will be discussed.

1.2.1 Computational prediction of IDPs

It has been reported that the amino acid sequences of IDPs are different from the sequences of ordered proteins.[15, 16, 64] IDPs tend to have both low content of hydrophobic and high content of similarly charged residues. In a more detailed comparison of amino acid sequences, it has been found that IDPs are depleted in order-promoting amino acids with aromatic or bulky aliphatic side-chains (Trp, Tye, Phe, Ile, Leu, and Val), and IDPs also have low content of Cys and Asn residues. On the other hand, IDPs tend to be enriched in disorder-promoting residues like Ala, Arg, Gly, Gln, Ser, Glu, and Lys, and also possess high content of the structure-breaking residue, Pro.[15, 35, 65] The differences between disordered and ordered proteins were not only limited to their amino acid composition, but can also be discriminated using some other physico-chemical properties, such as flexibility, hydropathy, β-sheet propensity, bulkiness, content of order and disorder-promoting residues, and net charge.[15] These properties became great resources to develop various computational tools to predict IDPs. There are about 70 predictors that are based on different principles and different computational approaches.[66] Among these predictors, the Predictor Of Natural Disordered Regions (PONDR) was one of the first developed predictors.[67, 68] PONDR is based on a sophisticated approach that uses algorithms trained to distinguish sequences that encode for disorder from those that encode for order. PONDR scores provide the propensity for amino acids to be disordered on a residue by residue basis. Residues with PONDR scores greater than 0.5 are predicted to be disordered, while scores below 0.5 are predicted to be ordered. Several versions of PONDR have been developed and one of its versions is specific for predicting short recognition elements is (VL-

XT). The PONDR-VL_XT algorithm [67, 69, 70] predictions for ACTR, CBP, and CaN are shown in Figure 1.3. The disorder prediction for ACTR indicates that all regions of ACTR are predicted to be disordered except for the middle region (residue 32-44) which is predicted to be slightly ordered. For CBP, the N-terminal region is predicted to be disordered, but the C-terminal region is predicted to be ordered. The prediction for CaN indicates that the regulatory domain is disordered except for a short region which is predicted to be ordered. The prediction for CaN correlates well with the structure of calcineurin determined by X-ray crystallography (see section 2.1 for the calcineurin domain structure).[62] Although disorder predictors can provide predictions based on amino acid sequences, these predictors cannot predict folding or unfolding that results from binding or solvent conditions. These weaknesses necessitate relying on experimental methods to study IDPs interactions and conformations.

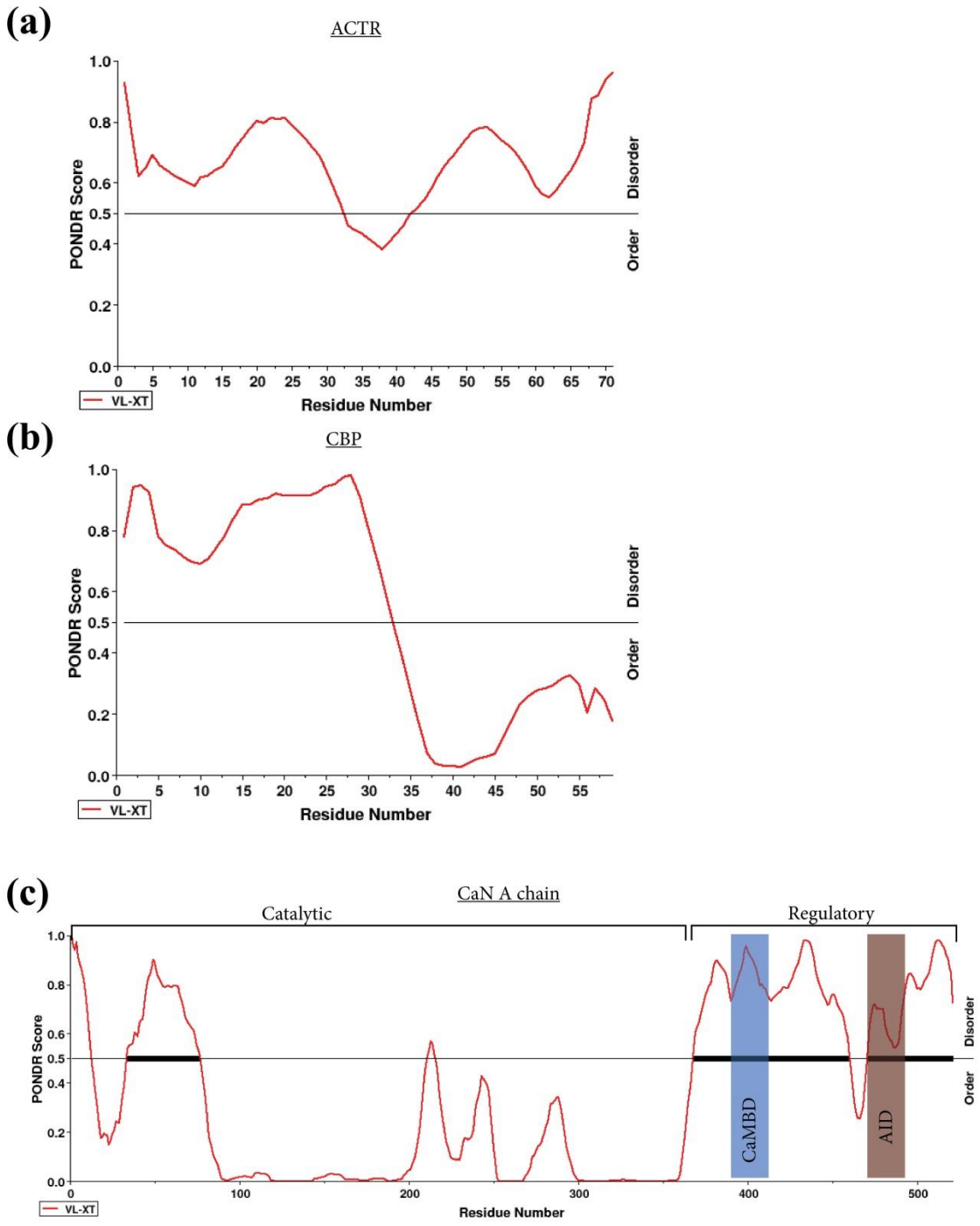


Figure 1.3. The PONDR-VL-XT disorder prediction for (a) ACTR, (b) CBP, and (c) calcineurin A chain. Residues with PONDR scores greater than 0.5 are predicted to be disordered and scores less than 0.5 are predicted to be ordered. The regulatory domain of calcineurin A chain is predicted to be disordered, which includes the calmodulin binding domain (CaMBD), blue shading and the autoinhibitory domain (AID), red shading.

1.2.2 Circular dichroism spectroscopy

One of the spectroscopic methods to identify IDPs is circular dichroism (CD). This method is based on measuring the difference of absorption of left-handed and right-handed circularly polarized light, which results from the chiral molecules.[52, 71, 72] Far-UV CD in the 190-230 nm region originates from optically active amide peptide bonds.[72, 73] It can be used to determine the relative amount of secondary structure content and also to distinguish α-helix, β-sheet, and random coil conformations. Near-UV CD in the range 250-350 nm region can also be a useful region to provide information on the tertiary structure because aromatic residues tend to have distinct wavelength profiles.[73, 74] CD has been a main technique for identification of IDPs and also for characterization of IDP structural changes induced by pH change, chemical denaturants, or temperature. Some examples of IDPs that were first classified as disordered protein based upon CD measurements are tau protein,[14] ProTa,[75] and α-synuclein.[13] One of the limitations of CD is that it cannot provide local structural information.

1.2.3 X-ray crystallography

The use of X-ray crystallography is not effective for analyzing fully disordered proteins due to these proteins' high conformational flexibility in their native states. The highly flexible nature makes these proteins unlikely to crystalize and even if they crystallize, the crystal state may not represent the native conformation.[76] However, the use of X-ray crystallography can be beneficial to analyze flexibility of regions within an ordered protein.[15] In fact, the crystallography method can be used as a reliable source of information on protein dynamics by analyzing the temperature-factor of atoms, also termed the B-factor, which can be taken as indicating the relative motion of different parts of the protein structure.[77] Atoms with high B-factors indicate that these atoms belong to part of the structure that is very flexible, and vice

versa. In addition, absence of electron density in protein structures very likely corresponds to disordered regions because increased mobility of atoms in these regions makes these atoms invisible by X-ray crystallography. There are a significant number of proteins in the DisProt database, a database for disordered proteins, having X-ray crystallography listed as their method of detection.[62, 78] One of the examples of an IDP that has a disordered region in its structure is calcineurin.[62] Calcineurin has a 95 residue region connecting two domains. Chapter 2 of this dissertation includes the structural study of this IDP in its free and complex forms.

1.2.4 Nuclear magnetic resonance spectroscopy

Nuclear Magnetic Resonance (NMR) spectroscopy is one of the most useful spectroscopic techniques to study IDPs in solution because it can provide information on the structure and dynamics with atomic resolution.[76] The structural information can be obtained from many parameters in NMR, such as chemical shift, relaxation rates, line widths, spin-spin multiplet patterns, and residual dipolar coupling.[77] NMR is based on the presence of nuclear spins and the magnetic moment of atomic nuclei like that of ^{1}H, ^{13}C, and ^{15}N.[52, 77] These nuclei have two possible spin states, which become split when an external magnetic field is applied. These nuclei can absorb electromagnetic radiation at a frequency that matches the energy difference between the two states. NMR can measure the resonance frequency and provide information on the local covalent and spatial arrangement of atoms, which depend on the chemical environment of resonating nuclei. NMR has been successfully applied to identify and characterize IDPs starting from the simplest NMR experiment, such as 1-D ^{1}H NMR, to provide global structural information such as in the case of gliotactin.[79] An example of 2-D NMR experiment is the heteronuclear single quantum coherence (HSQC)-based spectroscopy to provide residue-level information such as in the case of transcription factor ETS1.[80] With a 2-D

NMR experiment, the chemical shift of various nuclei can be measured. The deviation of chemical shift from complete random coil values, referred to as secondary chemical shift, for $^{1}H\alpha$, $^{13}C\alpha$, $^{13}\beta$, and ^{13}CO are sensitive to local secondary structure.[81] The random coil states can be experimentally obtained using different denaturing conditions like 8 M urea.[81] Other techniques have also been combined with NMR spectroscopy as a detection to study structure and function of IDPs. Since NMR can differentiate ^{1}H and ^{2}H nuclei, hydrogen exchange experiments in combination with NMR have been extensively used for the characterization of IDPs.[76, 82] In a recent paper, a Bayesian approach has been developed to determine the structural ensemble of IDPs by extracting valuable information from NMR chemical shifts and J-coupling data.[83] Although NMR has all the advantages mentioned above, NMR has also some limitations for IDP studies. One of the limitations is the high concentration requirement of the protein samples. Due to high structural flexibility, IDPs are more prone to aggregation at high concentrations.[82] One other limitation is that using NMR experiments for studying large proteins can be challenging due to NMR signal overlap.[84]

1.3 Disordered protein characterization by mass spectrometry

Mass spectrometry (MS) has emerged as a powerful tool for protein studies, which allows the investigation of the structure and the dynamics of proteins and their assemblies. The popularity of using MS evolved after the development of electrospray ionization (ESI)[85] and matrix assisted laser desorption/ionization (MALDI).[86] These ionization methods are considered "soft" ionization that can make ions out of biological molecules without causing fragmentation. Another advantage of ESI-MS is the ability to analyze protein mixture without prior purification steps as is required for other techniques, CD as an example. MS is also capable of analyzing large proteins, such as monoclonal antibodies[87, 88] and viral protein complexes.[89] In addition,

unlike NMR, MS can be utilized to analyze microliters of micromolar protein, which is of particular importance in the study of IDPs because IDPs are prone to aggregations at high concentrations.[82] Several MS-based techniques to study IDPs are discussed in the following sections. I discuss MS techniques based on preservation of noncovalent interactions under electrospray conditions, such as charge state distribution analysis and ion mobility. These two techniques work under native MS conditions, in which the native protein conformation can be preserved during the ESI process, termed native ESI.[90] The solvent for native ESI experiments should be compatible with the desolvation and ionization processes, and also should provide physiological conditions to preserve the protein native conformation. Solvent volatility and pH are two important properties for native ESI. Volatile buffer salts like ammonium acetate are usually used in the native ESI experiments.[91] I also discuss MS techniques based on solution-based labeling, such as hydrogen exchange and fast photochemical oxidation of proteins.

1.3.1 Charge state distribution analysis

For a pure protein, the typical MS spectrum consists of an envelope of signals where adjacent peaks represent ionization states differing by one charge, and the most intense peak is in the middle of the bell-shaped envelope. This envelope is referred to as the charge state distribution (CSD). Since CSDs are sensitive to the conformation held by proteins at the time of transfer to the gas phase, ESI-MS can be used to distinguish between ordered and disordered proteins by comparing their CSD profiles.[92] Proteins with a lower degree of compactness tend to have a wider CSD because the number of charges incorporated will be higher.[93, 94] Several factors need to be considered when designing ESI-MS experiments. These factors include solvent composition, pH, and effects of the source.[92] IDPs may have different conformational ensembles at different pHs. One examples of an IDP which has been studied by ESI-MS charge

state distribution analysis is α-synuclein.[95] α-synuclein is linked with Parkinson's disease due to α-synuclein's propensity to form intracellular aggregates.[96] It has been found that α-synuclein populates four different conformations which coexist in solution.[95] Each of these conformations has a different extent of disorder as identified by ESI-MS.

1.3.2 Ion mobility

One of the native ESI-MS techniques to study protein conformations is ion mobility (IM) coupled to ESI-MS. This technique allows distinguishing different conformers among ions of same mass/charge ratio (*m/z*) and can help reduce the spectral complexity of complicated protein mixtures in the gas phase.[94] The separation is based on a combination of cross-sectional area and *m/z*, in a chamber filled with a known gas (helium) at a known pressure across which is applied a weak static electric field.[97] The time required for the ions to travel through a cell is directly affected by the ion size and shape. Therefore, how quickly gas phase ions travel through the cell under the influence of the electric field can be measured and provide valuable information about the conformation and compactness of proteins.[97, 98] IM coupled to ESI-MS has been used to study several IDPs.[99-101] One of these IDPs is tumor suppressor protein, p53, which is known to bind to DNA.[102] The p53-DNA complex is stabilized by a zinc ion which plays a role in the p53 binding ability to DNA.[103] IM-MS has been used to investigate the conformational change of p53 caused by zinc.[104] Unfolded conformers have been detected in p53 in the absence of zinc when IM data are compared to the data of p53 in the absence of zinc. Although IM-MS can provide valuable information about IDPs, the technique still has some limitations. One of the limitations is in the mechanism of transferring ions from the solution phase to the gas phase for IDPs which may affect the actual native conformations of IDPs. IDPs become highly charged during ESI, which in turn increases the likelihood of forming extended

conformations due to electrostatic repulsion.[99] This electrostatic repulsion may disrupt any residual structure present within an IDP. Therefore, IM-MS may not be able to detect such residual structuring in IDPs in such cases. In addition, without complementary experiments, native ESI and IM-MS cannot provide localized information about folded or unfolded states in a protein.

1.3.3 Amide hydrogen exchange

Hydrogen exchange (HX) has become a powerful method to study protein conformation and dynamics. This method is based on monitoring the exchange by protein backbone amide hydrogens with deuterium atoms from the solvent surrounding the protein.[105, 106] The amide hydrogens exchange with deuterium at reaction rates that can be measured experimentally. Several techniques have been used to monitor hydrogen exchange kinetics in proteins. These techniques include infrared spectroscopy,[107] NMR,[108] and most recently mass spectrometry.[109] Hydrogen exchange coupled to mass spectrometry (HX-MS) has shown promise for analysis of proteins with high sensitivity, low sample requirement, and the ability to analyze large proteins and protein complexes.[110, 111] During the hydrogen exchange experiment, aliphatic hydrogens in proteins do not exchange to any measurable extent.[105] The deuterium exchange by side-chain (except for histidine) and N/C-terminal hydrogens will be very rapid, but the deuterium signature will not be retained in subsequent quenching and chromatography steps in a typical HX experiment.[105, 106, 112] The rate of amide HX is a function of the structure and dynamics of the protein because the exchange is highly dependent on hydrogen bonding and solvent accessibility.[113] Highly flexible regions will exchange more rapidly, while rigid regions exchange more slowly.[113, 114] Different HX experimental strategies have been used for investigating structure and dynamics in different protein and protein complexes. Using HX-MS,

proteins can either be studied intact or can be digested by proteolysis for local structural analysis. HX-MS have been frequently used to characterize IDPs. These IDPs include apolipoprotein A-I,[115] ACTR/CBP,[56] α-synuclein,[116] tau protein,[117] and methyl CpG-binding protein 2.[118] Since the research presented in the following two chapters utilizes HX-MS to study IDPs, a more in depth overview of HX-MS is presented next.

1.3.3.1 Fundamentals of hydrogen exchange

Amide HX in the absence of any higher order structure is referred to as chemical exchange. The chemical exchange rate constant (k_{ch}) for an amide can be expressed as follows:[119, 120]

$$k_{ch} = k_A[H_3O^+] + k_B[OH^-] + k_w[H_2O] \quad (1)$$

where k_A, k_B, and k_w are the rate constants for acid-, base-, and water-catalyzed amide deuterium exchange reaction, respectively. The rate constants, determined for a random coil-like poly-DL-alanine peptide at 20°C and low salt condition, are $k_A = 41.67\ M^{-1}\ min^{-1}$, $k_B = 1.12 \times 10^{10}\ M^{-1}\ min^{-1}$, and $k_A = 3.16 \times 10^{-2}\ M^{-1}\ min^{-1}$.[112] It is obvious that the rate of base catalysis exceeds the acid and water catalysis. Due to its low value, sometimes the water-catalyzed exchange is omitted from equation 1. The chemical exchange rate constant is highly dependent on pH, temperature, the identity of the amino acid residue and its neighboring side chains, and solution composition such as salt contents.[112] The chemical exchange rate is predominantly base-catalyzed above pH 3.0 and predominantly acid-catalyzed below pH 2.3. The rates of acid- and base-catalyzed exchange are equal at pH 2.5-3.0 (pH_{min}).[121] The chemical exchange rate increases ten-fold for every pH unit increase from pH_{min}. A temperature increase of 22 has a similar effect.[121]

Proteins in solution exists as an ensemble of interconverting conformations.[122] Therefore, each amide hydrogen in a protein may exist in folded and unfolded conformations at any point in time. The mechanism of HX usually described by the Linderstrom-Lang model.[105, 123] The mechanism is illustrated in equation 2.

$$NH_{closed} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} NH_{opened} \xrightarrow{k_{ch}} ND_{opened} \underset{k_{op}}{\overset{k_{cl}}{\rightleftharpoons}} ND_{closed} \qquad (2)$$

Each amide hydrogen is in dynamic equilibrium between two states: "closed" (exchange incompetent) and "open" (exchange competent). Breathing motions in proteins can transiently break hydrogen bonds in exchange incompetent state NH_{closed}, which exposes the exchangeable site to exchange with deuterium from the solvent at a rate defined by the chemical rate of exchange, k_{ch}.[124, 125] The frequency at which a residue will enter the exchange competent state NH_{open} will depend on the opening rate, k_{op}, before returning into the NH_{closed} state with the closing rate, k_{cl}. The observed hydrogen exchange rate for each residue, k_{obs}, which is the measured rate of deuterium loss or incorporation, can be expressed as follows:[126]

$$k_{obs} = \frac{k_{ch}k_{op}}{k_{cl}+k_{ch}+k_{op}} \qquad (3)$$

The equation 3 can be rewritten as follows if the amide is assumed to be mostly in the closed state, where $k_{op} << k_{cl}$:[126, 127]

$$k_{obs} \cong \frac{k_{ch}k_{op}}{k_{cl}+k_{ch}} \qquad (4)$$

The approximation in equation 4 always applies for folded proteins. Depending on the relative rates of the exchange and conformational transition in a protein, two distinct regimes, EX1 and

EX2, are commonly identified. Under EX2 conditions, $k_{cl} >> k_{ch}$ and the equation 4 can be rewritten as follows:[127]

$$k_{obs,EX2} = \frac{k_{op}}{k_{cl}} k_{ch} \qquad (5)$$

The kinetic limit under EX2 mechanism implies that the opening and closing of a hydrogen bond occurs many times before exchange actually occurs. From the equation 5, the protection factor, the ratio between chemical and observed exchange rates, can be measured. Measured kinetics of exchange under EX2 condition can offer a direct link to the free energy variation associated to local unfolding of a protein as described in equation 6:[126]

$$\Delta G_u^o = -RT \ln \frac{k_{op}}{k_{cl}} = -RT \ln \frac{k_{obs,EX2}}{k_{ch}} \qquad (6)$$

Where *R* is the gas constant and *T* is the absolute temperature. If the exchange is being monitored by MS, the characteristic of EX2 exchange will show a gradual increase in m/z of a peptide segment as a function of labeling time.[126]

The mechanism of exchange is referred to as EX1 when the structure closing step proceeds slower than the rate of exchange, $k_{cl} << k_{ch}$. Under EX1 condition the equation 4 can be re-written as follows:[127]

$$k_{obs,EX1} = k_{op} \qquad (7)$$

Under The EX1 limit, the deuterium exchange is much faster than conformational changes. There are few proteins that exchange by an EX1 mechanism under native conditions. If the exchange is being monitored by MS, the characteristic of EX1 exchange will show two distinct mass spectral envelopes on the m/z scale.[126] Two mass spectral envelopes will be observed

because $k_{op} << k_{cl}$ and $k_{cl} << k_{ch}$, which results in complete labeling of all exchangeable sites within a given protein at every opening event. As a function of labeling time, the intensity of the two envelopes changes, but not their positions on the m/z scale. In addition, as a function of time, a simultaneous decrease in the intensity of the unlabeled envelope and increase in the intensity of the labeled envelope will be observed.

1.3.3.2 HX experimental design and workflow

Protein HX reactions are typically performed at neutral or near-neutral pH. Mainly, there are two kinds of experimental schemes for HX labeling: pulse and continuous labeling.[(128)] In pulsed HX labeling, a protein is exposed to deuterium for a brief and constant time interval followed by transfer to quenching conditions (0 °C and pH around 2.5, where the chemical exchange rate is at minimum). Pulse labeling allows kinetic monitoring of folding and unfolding transitions.[(128)] For example, a folded protein can be introduced to a solution that has denatured promoting compositions. On the other hand, in continuous HX labeling, the protein is transferred to a deuterated solution for different lengths of time followed by transfer to quenching conditions. The labeling time for continuous labeling experiments can be various depending on the stability of intramolecular amide hydrogen bonds and solvent accessibility in a protein.[(126, 128)] Slowly exchanging regions will require longer exposure times to reach a deuterium maximum, but faster exchanging regions will require shorter exposure times to reach deuterium maxima. The rapid exchange of protein regions at neutral pH and 25 °C may indicate the presence of highly flexible or unstructured regions. Since continuous HX labeling uses different labeling time points (ranging from seconds to hours), the probability of detecting differences between two or more protein states is more likely. This property makes the continuous labeling experiments more widely applicable than the pulsed labeling experiments.

A schematic representation of a typical continuous HX-MS experimental workflow to analyze the local conformation of a protein in two states is shown in Figure 1.4. In the first step, the protein sample (panel A) is incubated in an excess amount of D_2O buffer for different amounts of time ranging from seconds to hours (standard time points) for a typical HX labeling experiments. Using the standard time points to study IDPs can be difficult because of several factors. One of the factors is the fast exchange that the majority of peptides will show in a largely unstructured protein, which makes these peptides reach deuterium maxima at the fastest time points of the standard labeling. Several developments have been made to the conventional HX experiment to probe the conformation and dynamics of IDPs. One of the methodological developments is the labeling of protein at millisecond time points.[129, 130] In the following section, an experimental design for millisecond HX experiment will be presented.

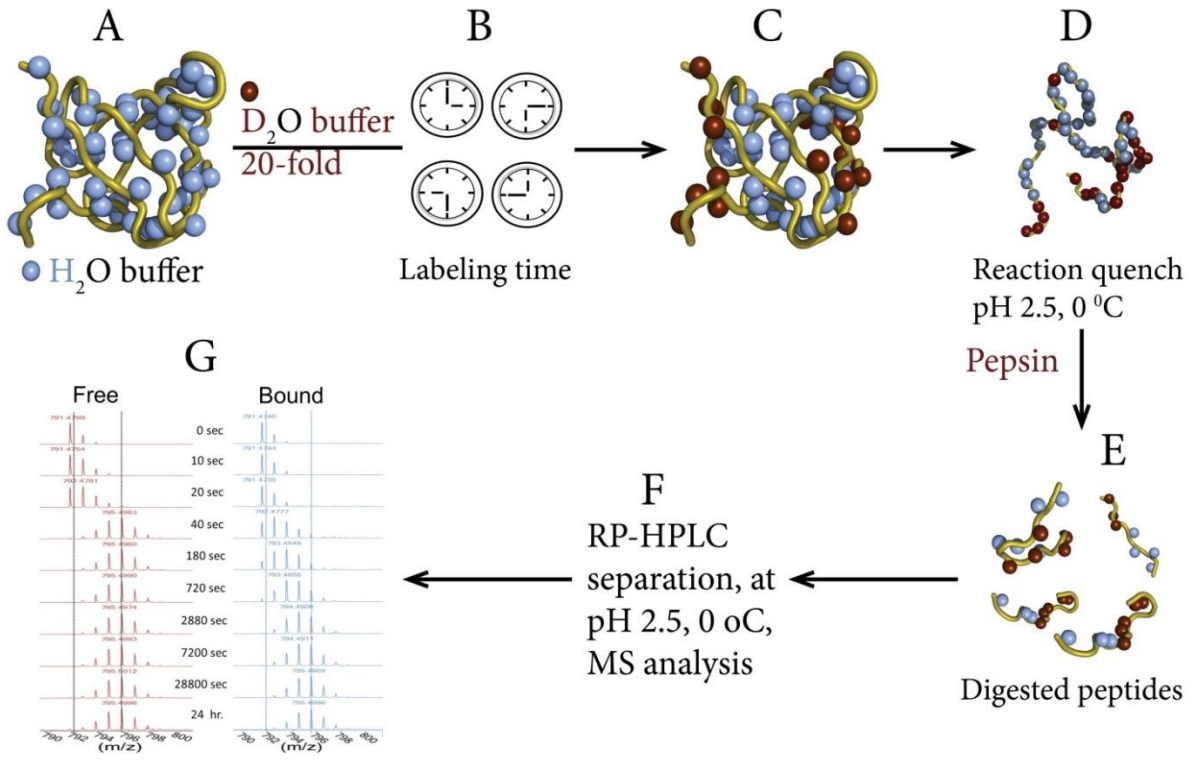


Figure 1.4. Schematic representation of continuous-labeling of the HX-MS experimental workflow. See the text for description of each step in the workflow.

In the next step of the HX experiment as shown in Figure 1.4, the labeled sample (panel C) is quenched by lowering the temperature to 0 °C and reducing the pH of the final mixture to 2.5. Reducing the pH can be achieved by adding of concentrated acid like hydrochloric acid, formic acid, or a strong buffer like phosphate. The rate of forward reaction (amide H to D) slows down by several orders of magnitude at quenching conditions.[112, 131] In the next step of Figure 1.4, to obtain the local structural information of at the peptide level, the quenched protein sample (panel D) is digested using an acid resistant protease like pepsin. The protein digestion can be achieved by either performing online pepsin digestion using an immobilized pepsin column[132, 133] or using in solution digestion with maintaining the quench conditions. The online digestion can offer a higher efficiency of digestion than in-solution digestion.[132] The digested peptides (panel E) can then be concentrated and desalted on a reversed-phase peptide trap. The peptides are separated using reversed-phase high performance liquid chromatography (HPLC). Then, the peptides are infused into a mass spectrometer through an ESI source for peptide mass analysis. Following data acquisition by the mass spectrometer, the identity of each peptide at different labeling time points and for each condition can be obtained (panel F). The deuterium incorporation at each time point can then be converted to an uptake plot to compare the kinetics of exchange for each peptide of the protein at a specific state with other states of the protein.

Several parameters can affect the total deuterium uptake for each peptide. One of the parameters is back exchange (D to H) during the quench step and subsequent chromatography steps.[109] During these steps, the H_2O based solution will add H_2O to the reaction mixture, which introduces new equilibrium conditions that leads to deuterium loss from the peptides. It is important to minimize back-exchange because higher deuterium loss causes loss of valuable information about protein dynamics. The labeling temperature,[131] pH,[112] buffer

compositions,[112] deuterium-to-sample ratio, and solution additives such as denaturants[134] during HX experiments play critical roles in controlling of back exchange. The steps to reduce back exchange include the use of refrigeration or ice baths to control temperature during chromatography, use of acid additives in the mobile phase to slow down the back exchange process, and the type of acid additive in the LC system. In order to determine the back exchange percentage in an HX experiment, a control experiment will be required. The control experiment, a fully deuterated control, can be prepared by labeling protein or peptides with D_2O for a much longer time than typical HX experiment. The control experiments will provide correction factors accounting for back exchange.[135]

1.3.3.3 Quench-flow apparatus for millisecond HX-MS

The labeling times in conventional HX experiments are in the range of seconds to hours, but using this time range to provide important information about the dynamic behavior of the most IDPs can be difficult. At physiological conditions, amide hydrogens in IDPs exchange at rates that are near to their chemical exchange rates. The rapid exchange of the IDP segments makes these peptides reach deuterium maxima at the fastest time points of the standard labeling time (usually around 10 s). Several developments have been made to the conventional HX experiment to characterize weakly structured regions of proteins or intact IDPs. Keppel and Weis developed an off-line quench-flow system for analysis of disordered proteins.[129] The labeling time is in the range between 42 to 5000 ms. A schematic depiction of the quench-flow system is shown in Figure 1.5. Three syringes deliver sample carrier (H_2O buffer, syringe 1 with 1 ml volume capacity), label (D_2O buffer, syringe 2 with 5 ml capacity), and quench (syringe 3 with 5 ml capacity). To maintain stable flow regardless of back pressure, all three syringes are driven by a single syringe pump. The connections are made with fused silica capillaries connected to the

three syringes with a special adapter to prevent leaks at high flow rates. The protein sample is introduced to the system using an electronically actuated six-port HPLC valve. The protein sample can be labeled with D_2O buffer at 1:5 (v:v) ratio at labeling mixer 1. To achieve a good mixing efficiency between the sample and D_2O buffer, a thinner fused silica internal diameter (ID) between the valve and T mixer 1 has been used. The two flows, sample (F_1) and label (F_2), are mixed at an angle of 180°. The resulting reaction mixture exits through mixer 1 and is labeled through a delay line with a known length and ID. The reaction time, t_{rxn}, has been determined from the mathematical relationship between the delay line volume, V_d (determined from delay line length, L, and inner radius, r_d), and total flow rate, F_1+F_2, as shown in equation 8:[(129)]

$$t_{\mathrm{rxn}} = \frac{V_{\mathrm{d}}}{F_1 + F_2} = \frac{\pi r_{\mathrm{d}}^2 L}{F_1 + F_2} \qquad (8)$$

By changing the flow rates, labeling times between 42 and 5250 ms have been achieved. The labeled sample is mixed with quench, delivered by syringe 3, in mixer 2 at an angle of 180°. The quenched sample exits mixer 2 and is immediately flash frozen in a microcentrifuge tube suspended in a bath of liquid nitrogen. The samples are kept at –80 °C until analysis by LC-MS system.

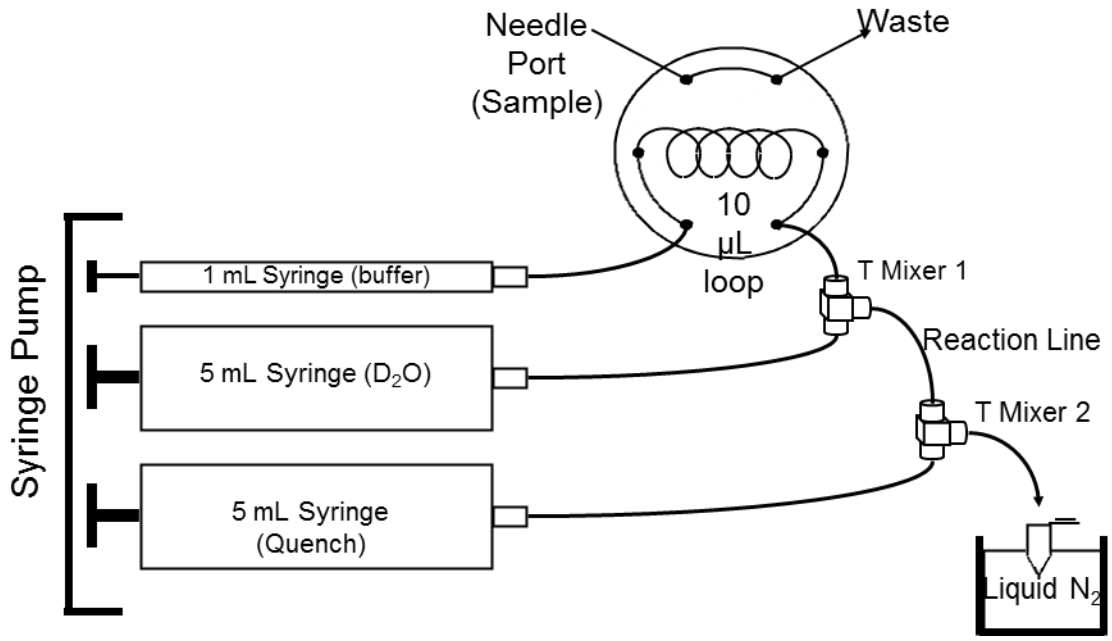


Figure 1.5. Schematic depiction of a setup for millisecond HX. See the text for more details on the setup overview. Redrawn from Ref. (129).

1.3.4 Hydroxyl radical protein footprinting

Another MS-based approach to investigate protein conformation and dynamics is hydroxyl radical protein footprinting. Hydroxyl radical, which can be generated by different methods, probes solvent accessibility of amino-acid side chains by oxidizing them.[(136)] The side chain modifications are irreversible and maintained throughout the following proteolysis and LC-MS/MS experiment.[(137)] These modifications result in mass shifts, which can be easily detected by MS and then by comparing the masses to unmodified forms of the protein. With hydroxyl radical footprinting, a change in solvent accessibility of a protein between two or more states can be determined.[(136, 137)] To localize the specific sites of oxidation, tandem mass spectrometry (MS/MS) methods should be used. Therefore, the structural resolution of hydroxyl radical methods is very high, and single side chain resolution can be achieved.[(136)]

1.3.4.1 Methods of generating hydroxyl radical

There are several methods for generating hydroxyl radicals to oxidize side chains. Detailed information about generating hydroxyl radical has been thoroughly reviewed.[138] One of the methods is Fenton chemistry.[139] This method is based on producing hydroxyl radical at slower rates, and this method can be used for experiments that use longer hydroxyl radical exposure times (5-30 s).[136, 139] The hydroxyl radicals in this method are produced from hydrogen peroxide by oxidizing EDTA-chelated Fe(II) to Fe(III). This method has become a standard method for studying DNA and RNA interactions because it allows for the stoichiometric generation of hydroxyl radical from Fe(II) and H_2O_2 without Fe (III) precipitation concerns. Another method of generating hydroxyl radical is synchrotron radiolysis of water. This method is based on radiolysis of water with X-ray photons.[140] As a result of the ejection of an electron from water, this high energy electron ionizes other water molecules, and ultimately generates hydroxyl radicals and other less reactive products. Generally, synchrotron radiolysis of water method generates a relatively high yield of hydroxyl radicals and can be used for experiments that use very short labeling times (10-200 ms).[136, 140] This method has been applied to footprinting DNA-protein interactions and large proteins. One of the advantages of this method is that there is no need for additives to produce hydroxyl radicals because water is the source of hydroxyl radical. Ultraviolet (UV) photolysis of hydrogen peroxide is another method to generate hydroxyl radicals.[141] UV light can homolytically cleave H_2O_2 to produce hydroxyl radicals. The hydroxyl radical exposure time using this method depends on the light source. It can be microseconds using laser sources or it can be as long as hours using UV lamps.[136] Although there are several methods to produce hydroxyl radicals, the results of oxidative chemical modifications are generally independent of the technique used for radical production.

While hydroxyl radicals are non-specific reactants, they do not label sidechains with equal efficiency. Hydroxyl radicals have been shown to react with 14 of the 20 naturally occurring amino acids.[136] These residues are Cys, Met, Trp, Tyr, Phe, His, Leu, Ile, Arg, Lys, Val, Ser, Gln, and Glu. The rate constant of the reactions range from 1.7×10^7 to 3.5×10^{10} $M^{-1}s^{-1}$.[136, 137] The hydroxyl radicals can also lead to nonspecific cleavage reactions of protein backbones, but this cleavage reaction occurs at rates 10-1000 times slower than side-chain oxidation.

1.3.4.2 Fast photochemical oxidation of proteins

Fast photochemical oxidation of protein (FPOP) is a method of hydroxyl radial footprinting that uses a pulsed KrF laser shot (248 nm) to generate hydroxyl radicals by photolysis of hydrogen peroxide (equation 9).[142, 143]

$$H_2O_2 \xrightarrow{h\nu} 2 \cdot OH \qquad (9)$$

The protein solution mixed with hydrogen peroxide and a dissolved hydroxyl scavenger like glutamine flows through a fused silica capillary. The protein sample is irradiated by a KrF excimer laser at a frequency that ensures all sample protein is irradiated once, but for a measurable exclusion fraction (see Figure 1.6). Figure 1.7 shows a schematic representation of FPOP experimental workflow for a protein hydroxyl radical labeling. Due to the presence of glutamine in the solution, the concentration of hydroxyl radical falls by 50,000-fold in under 1 microsecond, according to pseudo-first order analysis of hydroxyl radical consumption.[142] This timescale of labeling limits the oxidation to surface exposed side chains. This timescale is also faster than most protein unfolding events, so the labeling-perturbed conformations of protein are not sampled.[142, 143] The modified peptides can then be quantified using mass spectrometry after proteolysis of the modified protein sample. Although there are many pathways for hydroxyl

radical-mediated labeling on residue sidechains, the most common net mass shift is the increments of +15.9949 Da incorporation of oxygen in equilibrated solutions.[144] Different oxidation products of the peptides lead to a challenging task for data analysis. It is very difficult to do data analysis without using computational analytical tools. Several computational tools have been developed to meet the needs of covalent labeling experiments. One of the computational analytical tool is ProtMapMS.[145] The liquid chromatography elusion profile for each peptide can be automatically obtained using ProtMapMS. FPOP experiments have some advantages over HX experiments. One of the advantages in FPOP experiments is irreversible labeling of the peptides. Therefore, the FPOP experiments do not require specialized LC-MS-based strategies to preserve the label. Residue level resolution can be achieved by a tandem MS with FPOP experiments with no concern about gas-phase scrambling of the label, but label scrambling is problematic for HX experiments.[136] Since FPOP experiments probe only 14 out of the 20 amino acid side chain as mentioned above, FPOP experiments cannot be useful for proteins enriched in non-reactive amino acid residues. However, HX experiments can probe all amide backbone hydrogen, except proline. Since FPOP experiments probe side chains and HX experiments probe backbone amide hydrogens, the FPOP method can serve as a complementary method to HX in the conformation and dynamic study of proteins.

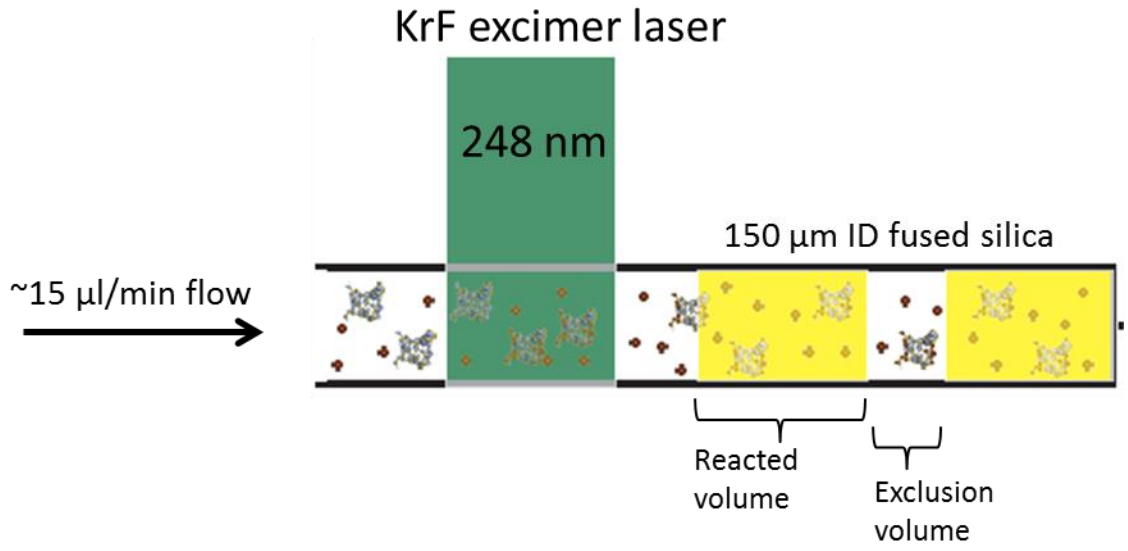


Figure 1.6. Schematic of the FPOP reaction region with typical settings. The laser beam intersects the fused silica flow tube (150 µm inner diameter). The frequency of laser beam and the flow rate are synchronized to insure that the protein is only exposed once to hydroxyl radicals. The synchronized flow and frequency of laser provide an exclusion region between two irradiated plugs.

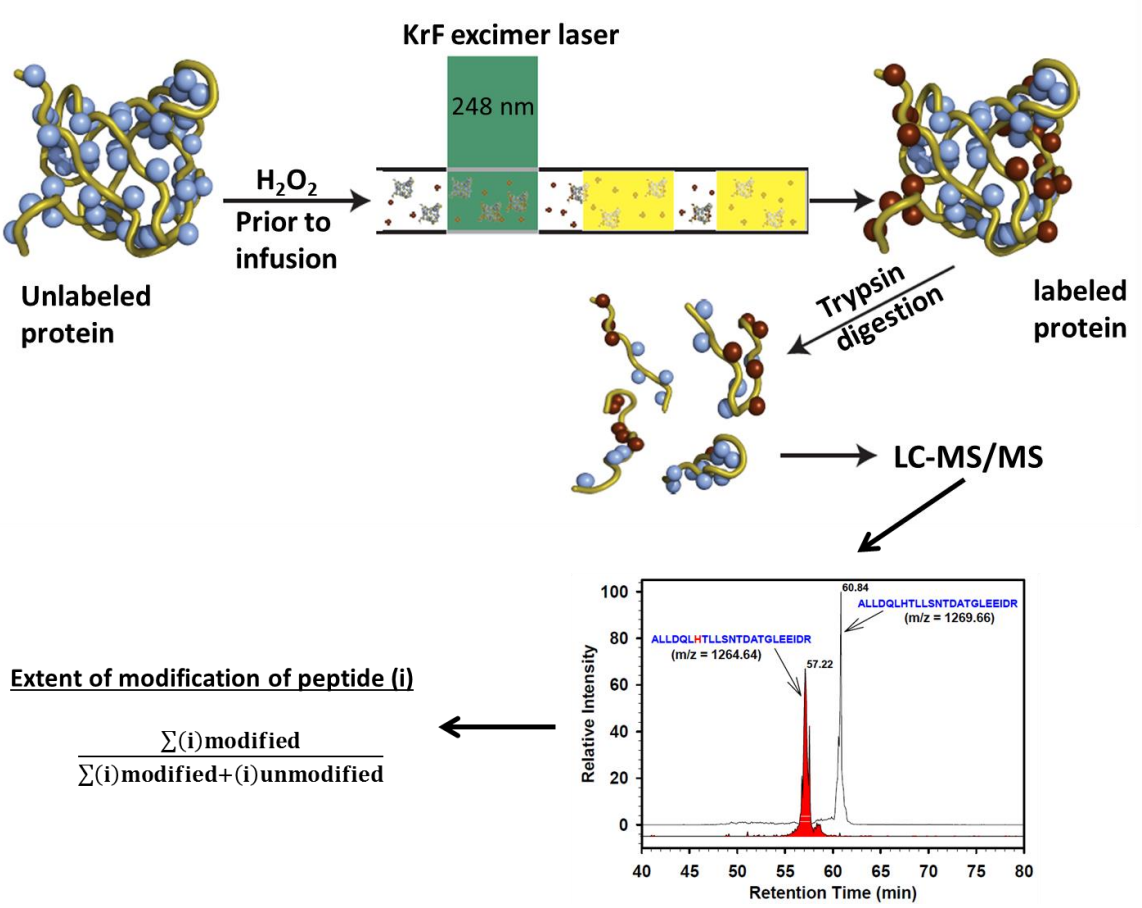


Figure 1.7. Schematic representation of the FPOP experimental workflow for a protein. In the first step, a protein sample is mixed with hydrogen peroxide and a hydroxyl radical scavenger like glutamine immediately prior to the infusion. In the next step, the protein sample is labeled with hydroxyl radicals that are generated by pulsed KrF excimer laser shots (248 nm). The next step is protein digestion using trypsin. The tryptic peptides are then injected into the LC-MS/MS system, and then analyzed by a mass spectrometry. The data obtained from LC-MS/MS contains the total ion chromatograms, extracted ion chromatogram, and mass spectra for all peptides. In the final step, the fraction of modification for each peptide can be calculated, and residue-level resolution can be achieved by MS/MS data analysis for each peptide.

1.4 Overview of the dissertation

Due to important roles of IDPs in cells, it is essential to develop new methods to study conformational dynamics of IDPs. Over the past two decades mass spectrometry has become a powerful technique for biophysical characterization of proteins due to its high sensitivity and variety of choices for sample preparation and instrumentation. The purpose of this dissertation is

to demonstrate the application of MS-based approaches to study IDPs. Here I present the application of two important mass spectrometry-based approaches: hydrogen exchange and fast photochemical oxidation of protein to characterize two disordered protein systems calcineurin and ACTR/CBP, respectively. This research was also conducted to make improvements to HX-MS methods for IDPs studies. The following chapters discuss the advantages of using these two MS-based techniques in the applications of complicated IDP systems.

Calcineurin plays essential roles in processes such as T-cell activation, neuronal and muscle development, and cardiac hypertrophy. In its inactive state, an autoinhibitory domain is bound in the active site. At high calcium concentration, calmodulin binds calcium ions, resulting in calmodulin binding to the unstructured regulatory domain of calcineurin. Calmodulin binding causes release of the autoinhibitory domain from the active site, activating calcineurin. To better understand the mechanism of activation, it is important to look at the complete complex. To identify conformational and dynamic changes in calcineurin induced by calmodulin, HX-MS has been used. The HX-MS data for calcineurin is discussed in the chapter 2. The results are presented from all regions of the CaN A and B chains to describe the mechanism of CaN activation. A new model of CaN activation upon CaM binding is discussed.

Millisecond HX can be a very useful technique to characterize IDPs because most of the exchange occurs on this timescale at room temperature and near neutral pH. Protection in HX experiments, which is a measure of the slowing of exchange caused by structure, can reveal information about protein structures. The degree of protection can be directly related to the degree of structure if the observed rate of HX is compared to the rate of HX by an amide in an unprotected reference state. Without an unprotected reference state, it is impossible to estimate such protection because the unprotected reference state is an essential part in measuring

protection. Generally, chemical exchange, calculated using an empirical formula based on unstructured model peptides, is used as an unprotected reference state. Our lab and others have found that the calculated chemical exchange does not seem to be accurate enough to be used as an unprotected reference state in weakly protected IDPs. In the chapter 3, different approaches to prepare an unprotected reference state for quantifying the degree of structure in IDPs are explored. ACTR and CBP were used as model IDPs to explore the best approach to produce an unprotected reference state for millisecond HX.

ACTR and CBP are a useful model system for conformational analysis of IDPs. On their own, CBP is a mostly helical molten globule while ACTR is near-random coil. The proteins bind with high affinity to form a well-defined complex with three helices each. Our previous work showed that HX can provide important information about IDPs, but labeling takes place on the millisecond-second timescale. Since protein folding is slower than microseconds, FPOP might reveal states of IDPs that are undetectable by HX. Chapter 4 discusses the application of FPOP method to characterize IDPs. This chapter also discusses how FPOP method can serve as a complementary method to HX in the study of localized disorder regions in IDPs.

1.5 References

1. Lemieux, R. U., and Spohr, U. (1994) How Emil Fischer was led to the lock and key concept for enzyme specificity, In *Advances in carbohydrate chemistry and biochemistry* (Derek, H., Ed.), pp 1-20, Academic Press, Cambridge, Massachusetts.

2. Fischer, E. (1894) Einfluss der configuration auf die wirkung der enzyme, *Berichte der deutschen chemischen Gesellschaft 27*, 2985-2993.

3. Uversky, V. N. (2014) Intrinsically disordered proteins, In *Protein folding and structure* (Gomes, C. M., Ed.), pp 1-61, Springer Cham Heidelberg New York Dordrecht, London.

4. Uversky, V. N. (2013) A decade and a half of protein intrinsic disorder: biology still waits for physics, *Protein science 22*, 693-724.

5. Uversky, V. N., and Dunker, A. K. (2010) Understanding protein non-folding, *Biochimica et biophysica acta 1804*, 1231-1264.

6. Huber, R., and Bennett, W. S., Jr. (1983) Functional significance of flexibility in proteins, *Biopolymers 22*, 261-279.

7. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nature reviews. Molecular cell biology 6*, 197-208.

8. Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Obradovic, Z., and Uversky, V. N. (2007) Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins, *Journal of proteome research 6*, 1917-1932.

9. Bae, S. H., Dyson, H. J., and Wright, P. E. (2009) Prediction of the rotational tumbling time for proteins with disordered segments, *Journal of the American chemical society 131*, 6814-6821.

10. Tantos, A., Han, K. H., and Tompa, P. (2012) Intrinsic disorder in cell signaling and gene transcription, *Molecular and cellular endocrinology 348*, 457-465.

11. Linderstrom-Lang, K. S., J. A. (1959) protein structure and enzyme activity, In *The Enzymes; 2nd ed.; (Boyer, P. D., Lardy, H., Myrback, K., Eds.);* , Academic Press: New York.

12. Daughdrill, G. W. P., G. J.; Uversky, V. N.; Cortese, M. S.; Dunker, A. K. (2005) *Handbook of protein folding, (Buchner, J., Kiefhaber, T., Eds.);*, Wiley-VCH, Verlag GmbH & Co. KGaA: Weinheim, Germany.

13. Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A., and Lansbury, P. T., Jr. (1996) NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded, *Biochemistry 35*, 13709-13715.

14. Schweers, O., Schonbrunn-Hanebeck, E., Marx, A., and Mandelkow, E. (1994) Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure, *The Journal of biological chemistry 269*, 24290-24297.

15. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001) Intrinsically disordered protein, *Journal of molecular graphics & modelling 19*, 26-59.

16. Dunker, A. K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C., and Villafranca, J. E. (1998) Protein disorder and the evolution of molecular recognition: theory, predictions and observations, *Pacific symposium on biocomputing*, 473-484.

17. Garner, E., Cannon, P., Romero, P., Obradovic, Z., and Dunker, A. K. (1998) Predicting disordered regions from amino acid sequence: Common themes despite differing structural characterization, *Genome informatics. Workshop on genome informatics 9*, 201-213.

18. van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E., and Babu, M. M. (2014) Classification of intrinsically disordered regions and proteins, *Chemical reviews 114*, 6589-6631.

19. Hilser, V. J., Garcia-Moreno, E. B., Oas, T. G., Kapp, G., and Whitten, S. T. (2006) A statistical thermodynamic model of the protein ensemble, *Chemical reviews 106*, 1545-1558.

20. Hughson, F. M., Wright, P. E., and Baldwin, R. L. (1990) Structural characterization of a partly folded apomyoglobin intermediate, *Science 249*, 1544-1548.

21. Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J., Evans, R. M., and Wright, P. E. (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators, *Nature 415*, 549-553.

22. Ebert, M. O., Bae, S. H., Dyson, H. J., and Wright, P. E. (2008) NMR relaxation study of the complex formed between CBP and the activation domain of the nuclear hormone receptor coactivator ACTR, *Biochemistry 47*, 1299-1308.

23. Uversky, V. N., and Ptitsyn, O. B. (1994) "Partly folded" state, a new equilibrium state of protein molecules: four-state guanidinium chloride-induced unfolding of beta-lactamase at low temperature, *Biochemistry 33*, 2782-2791.

24. Uversky, V. N., and Ptitsyn, O. B. (1996) Further evidence on the equilibrium "pre-molten globule state": four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature, *Journal of molecular biology 255*, 215-228.

25. Shortle, D., and Ackerman, M. S. (2001) Persistence of native-like topology in a denatured protein in 8 M urea, *Science 293*, 487-489.

26. Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C., and Brown, C. J. (2000) Intrinsic protein disorder in complete genomes, *Genome informatics. Workshop on genome informatics 11*, 161-171.

27. Ferron, F., Longhi, S., Canard, B., and Karlin, D. (2006) A practical overview of protein disorder prediction methods, *Proteins 65*, 1-14.

28. Dosztanyi, Z., Sandor, M., Tompa, P., and Simon, I. (2007) Prediction of protein disorder at the domain level, *Current protein & peptide science 8*, 161-171.

29. He, B., Wang, K., Liu, Y., Xue, B., Uversky, V. N., and Dunker, A. K. (2009) Predicting intrinsic disorder in proteins: an overview, *Cell research 19*, 929-949.

30. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life, *Journal of molecular biology 337*, 635-645.

31. Tompa, P., Dosztanyi, Z., and Simon, I. (2006) Prevalent structural disorder in E. coli and S. cerevisiae proteomes, *Journal of proteome research 5*, 1996-2000.

32. Galea, C. A., High, A. A., Obenauer, J. C., Mishra, A., Park, C. G., Punta, M., Schlessinger, A., Ma, J., Rost, B., Slaughter, C. A., and Kriwacki, R. W. (2009) Large-scale analysis of thermostable, mammalian proteins provides insights into the intrinsically disordered proteome, *Journal of proteome research 8*, 211-226.

33. Oates, M. E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M. J., Xue, B., Dosztanyi, Z., Uversky, V. N., Obradovic, Z., Kurgan, L., Dunker, A. K., and Gough, J. (2013) D(2)P(2): database of disordered protein predictions, *Nucleic acids research 41*, D508-516.

34. Burger, V. M., Nolasco, D. O., and Stultz, C. M. (2016) Expanding the Range of Protein Function at the Far End of the Order-Structure Continuum, *The Journal of biological chemistry 291*, 6706-6713.

35. Radivojac, P., Iakoucheva, L. M., Oldfield, C. J., Obradovic, Z., Uversky, V. N., and Dunker, A. K. (2007) Intrinsic disorder and functional proteomics, *Biophysical journal 92*, 1439-1456.

36. Kovacs, D., Szabo, B., Pancsa, R., and Tompa, P. (2013) Intrinsically disordered proteins undergo and assist folding transitions in the proteome, *Archives of biochemistry and biophysics 531*, 80-89.

37. Uversky, V. N., and Fink, A. L. (2004) Conformational constraints for amyloid fibrillation: the importance of being unfolded, *Biochimica et biophysica acta 1698*, 131-153.

38. Masters, C. L., Multhaup, G., Simms, G., Pottgiesser, J., Martins, R. N., and Beyreuther, K. (1985) Neuronal origin of a cerebral amyloid: neurofibrillary tangles of Alzheimer's disease contain the same protein as the amyloid of plaque cores and blood vessels, *The EMBO journal 4*, 2757-2763.

39. Lee, V. M., Balin, B. J., Otvos, L., Jr., and Trojanowski, J. Q. (1991) A68: a major subunit of paired helical filaments and derivatized forms of normal Tau, *Science 251*, 675-678.

40. Ueda, K., Fukushima, H., Masliah, E., Xia, Y., Iwai, A., Yoshimoto, M., Otero, D. A., Kondo, J., Ihara, Y., and Saitoh, T. (1993) Molecular cloning of cDNA encoding an

unrecognized component of amyloid in Alzheimer disease, *Proceedings of the national academy of sciences of the United States of America 90*, 11282-11286.

41. Lee, H., Mok, K. H., Muhandiram, R., Park, K. H., Suk, J. E., Kim, D. H., Chang, J., Sung, Y. C., Choi, K. Y., and Han, K. H. (2000) Local structural elements in the mostly unstructured transcriptional activation domain of human p53, *The Journal of biological chemistry 275*, 29426-29432.

42. Chang, B. S., Minn, A. J., Muchmore, S. W., Fesik, S. W., and Thompson, C. B. (1997) Identification of a novel regulatory domain in Bcl-X(L) and Bcl-2, *The EMBO journal 16*, 968-977.

43. Campbell, K. M., Terrell, A. R., Laybourn, P. J., and Lumb, K. J. (2000) Intrinsic structural disorder of the C-terminal activation domain from the bZIP transcription factor Fos, *Biochemistry 39*, 2708-2713.

44. Sunde, M., McGrath, K. C., Young, L., Matthews, J. M., Chua, E. L., Mackay, J. P., and Death, A. K. (2004) TC-1 is a novel tumorigenic and natively disordered protein associated with thyroid cancer, *Cancer research 64*, 2766-2773.

45. Glenner, G. G., and Wong, C. W. (1984) Alzheimer's disease and Down's syndrome: sharing of a unique cerebrovascular amyloid fibril protein, *Biochemical and biophysical research communications 122*, 1131-1135.

46. Wisniewski, K. E., Dalton, A. J., McLachlan, C., Wen, G. Y., and Wisniewski, H. M. (1985) Alzheimer's disease in Down's syndrome: clinicopathologic studies, *Neurology 35*, 957-961.

47. Dev, K. K., Hofele, K., Barbieri, S., Buchman, V. L., and van der Putten, H. (2003) Part II: alpha-synuclein and its molecular pathophysiological role in neurodegenerative disease, *Neuropharmacology 45*, 14-44.

48. Wright, P. E., and Dyson, H. J. (2009) Linking folding and binding, *Current opinion in structural biology 19*, 31-38.

49. Lacy, E. R., Filippov, I., Lewis, W. S., Otieno, S., Xiao, L., Weiss, S., Hengst, L., and Kriwacki, R. W. (2004) p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding, *Nature structural & molecular biology 11*, 358-364.

50. Shammas, S. L., Crabtree, M. D., Dahal, L., Wicky, B. I., and Clarke, J. (2016) Insights into Coupled Folding and Binding Mechanisms from Kinetic Studies, *The Journal of biological chemistry 291*, 6689-6695.

51. Uversky, V. N. (2011) Intrinsically disordered proteins from A to Z, *The International journal of biochemistry & cell biology 43*, 1090-1103.

52. Receveur-Brechot, V., Bourhis, J. M., Uversky, V. N., Canard, B., and Longhi, S. (2006) Assessing protein disorder and induced folding, *Proteins 62*, 24-45.

53. Arai, M., Sugase, K., Dyson, H. J., and Wright, P. E. (2015) Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding, *Proceedings of the national academy of sciences of the United States of America 112*, 9614-9619.

54. Kjaergaard, M., Norholm, A. B., Hendus-Altenburger, R., Pedersen, S. F., Poulsen, F. M., and Kragelund, B. B. (2010) Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II?, *Protein science : a publication of the protein society 19*, 1555-1564.

55. Kjaergaard, M., Iešmantavičius, V., and Poulsen, F. M. (2011) The interplay between transient α-helix formation and side chain rotamer distributions in disordered proteins probed by methyl chemical shifts, *Protein science : a publication of the protein society 20*, 2023-2034.

56. Keppel, T. R., Howard, B. A., and Weis, D. D. (2011) Mapping Unstructured Regions and Synergistic Folding in Intrinsically Disordered Proteins with Amide H/D Exchange Mass Spectrometry, *Biochemistry 50*, 8722-8732.

57. Iesmantavicius, V., Jensen, M. R., Ozenne, V., Blackledge, M., Poulsen, F. M., and Kjaergaard, M. (2013) Modulation of the intrinsic helix propensity of an intrinsically disordered protein reveals long-range helix-helix interactions, *Journal of the American Chemical Society 135*, 10155-10163.

58. Wang, J. H., and Desai, R. (1976) A brain protein and its effect on the Ca2+-and protein modulator-activated cyclic nucleotide phosphodiesterase, *Biochemical and biophysical research communications 72*, 926-932.

59. Aramburu, J., Rao, A., and Klee, C. B. (2000) Calcineurin: from structure to function, *Current topics in cellular regulation 36*, 237-295.

60. Rusnak, F., and Mertz, P. (2000) Calcineurin: form and function, *Physiological reviews 80*, 1483-1521.

61. Li, H., Rao, A., and Hogan, P. G. (2011) Interaction of calcineurin with substrates and targeting proteins, *Trends in cell biology 21*, 91-103.

62. Kissinger, C. R., Parge, H. E., Knighton, D. R., Lewis, C. T., Pelletier, L. A., Tempczyk, A., Kalish, V. J., Tucker, K. D., Showalter, R. E., Moomaw, E. W., and et al. (1995) Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex, *Nature 378*, 641-644.

63. Klee, C. B., Ren, H., and Wang, X. (1998) Regulation of the calmodulin-stimulated protein phosphatase, calcineurin, *The Journal of biological chemistry 273*, 13367-13370.

64. Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions?, *Proteins 41*, 415-427.

65. Campen, A., Williams, R. M., Brown, C. J., Meng, J., Uversky, V. N., and Dunker, A. K. (2008) TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder, *Protein and peptide letters 15*, 956-963.

66. Li, J., Feng, Y., Wang, X., Li, J., Liu, W., Rong, L., and Bao, J. (2015) An Overview of Predictors for Intrinsically Disordered Proteins over 2010-2014, *International journal of molecular sciences 16*, 23446-23462.

67. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein, *Proteins 42*, 38-48.

68. Williams, R. M., Obradovi, Z., Mathura, V., Braun, W., Garner, E. C., Young, J., Takayama, S., Brown, C. J., and Dunker, A. K. (2001) The protein non-folding problem: amino acid determinants of intrinsic order and disorder, *Pacific symposium on biocomputing*, 89-100.

69. Romero, Obradovic, and Dunker, K. (1997) Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family, *Genome informatics. Workshop on Genome Informatics 8*, 110-124.

70. Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999) Predicting Protein Disorder for N-, C-, and Internal Regions, *Genome informatics. Workshop on genome informatics 10*, 30-40.

71. Nordén, B., Rodger, A., Dafforn, T., and Chemistry, R. S. o. (2010) *Linear Dichroism and Circular Dichroism: A textbook on polarized-light spectroscopy*, pp 1-293, Royal Society of Chemistry, Piccadilly, London.

72. Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure, *Nature protocols 1*, 2876-2890.

73. Whitmore, L., and Wallace, B. A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases, *Biopolymers 89*, 392-400.

74. Chou, D. K., Krishnamurthy, R., Manning, M. C., Randolph, T. W., and Carpenter, J. F. (2012) Physical stability of albinterferon-alpha(2b) in aqueous solution: effects of conformational stability and colloidal stability on aggregation, *Journal of pharmaceutical sciences 101*, 2702-2719.

75. Gast, K., Damaschun, H., Eckert, K., Schulze-Forster, K., Maurer, H. R., Muller-Frohne, M., Zirwer, D., Czarnecki, J., and Damaschun, G. (1995) Prothymosin alpha: a biologically active protein with random coil conformation, *Biochemistry 34*, 13211-13218.

76. Dyson, H. J., and Wright, P. E. (2004) Unfolded proteins and protein folding studied by NMR, *Chemical reviews 104*, 3607-3622.

77. Tompa, P., and Fersht, A. (2009) *Structure and function of intrinsically disordered proteins*, Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton, FL.

78. Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution, *Science 292*, 1863-1876.

79. Zeev-Ben-Mordehai, T., Rydberg, E. H., Solomon, A., Toker, L., Auld, V. J., Silman, I., Botti, S., and Sussman, J. L. (2003) The intracellular domain of the Drosophila cholinesterase-like neural adhesion protein, gliotactin, is natively unfolded, *Proteins 53*, 758-767.

80. Macauley, M. S., Errington, W. J., Scharpf, M., Mackereth, C. D., Blaszczak, A. G., Graves, B. J., and McIntosh, L. P. (2006) Beads-on-a-string, characterization of ETS-1 sumoylated within its flexible N-terminal sequence, *The Journal of biological chemistry 281*, 4164-4172.

81. Schwarzinger, S., Kroon, G. J., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) Sequence-dependent correction of random coil NMR chemical shifts, *Journal of the American Chemical Society 123*, 2970-2978.

82. Sharma, S., Zheng, H., Huang, Y. J., Ertekin, A., Hamuro, Y., Rossi, P., Tejero, R., Acton, T. B., Xiao, R., Jiang, M., Zhao, L., Ma, L. C., Swapna, G. V., Aramini, J. M., and Montelione, G. T. (2009) Construct optimization for protein NMR structure analysis using amide hydrogen/deuterium exchange mass spectrometry, *Proteins 76*, 882-894.

83. Brookes, D. H., and Head-Gordon, T. (2016) Experimental Inferential Structure Determination of Ensembles for Intrinsically Disordered Proteins, *Journal of the American Chemical Society 138*, 4530-4538.

84. Kaltashov, I. A., Bobst, C. E., Abzalimov, R. R., Berkowitz, S. A., and Houde, D. (2010) Conformation and dynamics of biopharmaceuticals: transition of mass spectrometry-based tools from academe to industry, *Journal of the American society for mass spectrometry 21*, 323-337.

85. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules, *Science 246*, 64-71.

86. Karas, M., and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons, *Analytical chemistry 60*, 2299-2301.

87. Bondarenko, P. V., Second, T. P., Zabrouskov, V., Makarov, A. A., and Zhang, Z. (2009) Mass measurement and top-down HPLC/MS analysis of intact monoclonal antibodies on a hybrid linear quadrupole ion trap-Orbitrap mass spectrometer, *Journal of the American society for mass spectrometry 20*, 1415-1424.

88. Du, Y., Wang, F., May, K., Xu, W., and Liu, H. (2012) LC-MS analysis of glycopeptides of recombinant monoclonal antibodies by a rapid digestion procedure, *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences 907*, 87-93.

89. Eyckerman, S., Titeca, K., Van Quickelberghe, E., Cloots, E., Verhee, A., Samyn, N., De Ceuninck, L., Timmerman, E., De Sutter, D., Lievens, S., Van Calenbergh, S., Gevaert, K., and Tavernier, J. (2016) Trapping mammalian protein complexes in viral particles, *Nature communications 7*, 11416.

90. Sharon, M., and Robinson, C. V. (2007) The role of mass spectrometry in structure elucidation of dynamic protein complexes, *Annual review of biochemistry 76*, 167-193.

91. Heck, A. J. (2008) Native mass spectrometry: a bridge between interactomics and structural biology, *Nature methods 5*, 927-933.

92. Kaltashov, I. A., and Abzalimov, R. R. (2008) Do ionic charges in ESI MS provide useful information on macromolecular structure?, *Journal of the American society for mass spectrometry 19*, 1239-1246.

93. Konermann, L., and Douglas, D. J. (1997) Acid-induced unfolding of cytochrome c at different methanol concentrations: electrospray ionization mass spectrometry specifically monitors changes in the tertiary structure, *Biochemistry 36*, 12296-12302.

94. Vis, H., Heinemann, U., Dobson, C. M., and Robinson, C. (1998) Detection of a monomeric intermediate associated with dimerization of protein HU by mass spectrometry, *Journal of the American Chemical Society 120*, 6427-6428.

95. Frimpong, A. K., Abzalimov, R. R., Uversky, V. N., and Kaltashov, I. A. (2010) Characterization of intrinsically disordered proteins with electrospray ionization mass spectrometry: conformational heterogeneity of alpha-synuclein, *Proteins 78*, 714-722.

96. Baba, M., Nakajo, S., Tu, P. H., Tomita, T., Nakaya, K., Lee, V. M., Trojanowski, J. Q., and Iwatsubo, T. (1998) Aggregation of alpha-synuclein in Lewy bodies of sporadic Parkinson's disease and dementia with Lewy bodies, *The American journal of pathology 152*, 879-884.

97. Pacholarz, K. J., Garlish, R. A., Taylor, R. J., and Barran, P. E. (2012) Mass spectrometry based tools to investigate protein-ligand interactions for drug discovery, *Chemical society reviews 41*, 4335-4355.

98. Beveridge, R., Chappuis, Q., Macphee, C., and Barran, P. (2013) Mass spectrometry methods for intrinsically disordered proteins, *The Analyst 138*, 32-42.

99. Brocca, S., Testa, L., Sobott, F., Samalikova, M., Natalello, A., Papaleo, E., Lotti, M., De Gioia, L., Doglia, S. M., Alberghina, L., and Grandori, R. (2011) Compaction properties of an intrinsically disordered protein: Sic1 and its kinase-inhibitor domain, *Biophysical journal 100*, 2243-2252.

100. Canon, F., Ballivian, R., Chirot, F., Antoine, R., Sarni-Manchado, P., Lemoine, J., and Dugourd, P. (2011) Folding of a salivary intrinsically disordered protein upon binding to tannins, *Journal of the American Chemical Society 133*, 7847-7852.

101. Bernstein, S. L., Liu, D., Wyttenbach, T., Bowers, M. T., Lee, J. C., Gray, H. B., and Winkler, J. R. (2004) Alpha-synuclein: stable compact and extended monomeric structures and pH dependence of dimer formation, *Journal of the American Society for Mass Spectrometry 15*, 1435-1443.

102. Hainaut, P., and Mann, K. (2001) Zinc binding and redox control of p53 structure and function, *Antioxidants & redox signaling 3*, 611-623.

103. Meplan, C., Richard, M. J., and Hainaut, P. (2000) Metalloregulation of the tumor suppressor protein p53: zinc mediates the renaturation of p53 after exposure to metal chelators in vitro and in intact cells, *Oncogene 19*, 5227-5236.

104. Faull, P. A., Florance, H. V., Schmidt, C. Q., Tomczyk, N., Barlow, P. N., Hupp, T. R., Nikolova, P. V., and Barran, P. E. (2010) Utilising ion mobility-mass spectrometry to interrogate macromolecules: Factor H complement control protein modules 10–15 and 19–20 and the DNA-binding core domain of tumour suppressor p53, *International journal of mass spectrometry 298*, 99-110.

105. Konermann, L., Pan, J., and Liu, Y. H. (2011) Hydrogen exchange mass spectrometry for studying protein structure and dynamics, *Chemical society reviews 40*, 1224-1234.

106. Iacob, R. E., and Engen, J. R. (2012) Hydrogen exchange mass spectrometry: are we out of the quicksand?, *Journal of the American society for mass spectrometry 23*, 1003-1010.

107. Kim, Y. S., Randolph, T. W., Manning, M. C., Stevens, F. J., and Carpenter, J. F. (2003) Congo red populates partially unfolded states of an amyloidogenic protein to enhance aggregation and amyloid fibril formation, *The Journal of biological chemistry 278*, 10842-10850.

108. Paterson, Y., Englander, S. W., and Roder, H. (1990) An antibody binding site on cytochrome c defined by hydrogen exchange and two-dimensional NMR, *Science 249*, 755-759.

109. Zhang, Z., and Smith, D. L. (1993) Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation, *Protein science : a publication of the Protein Society 2*, 522-531.

110. Houde, D., Peng, Y., Berkowitz, S. A., and Engen, J. R. (2010) Post-translational modifications differentially affect IgG1 conformation and receptor binding, *Molecular & cellular proteomics : MCP 9*, 1716-1728.

111. Houde, D., Arndt, J., Domeier, W., Berkowitz, S., and Engen, J. R. (2009) Characterization of IgG1 Conformation and Conformational Dynamics by Hydrogen/Deuterium Exchange Mass Spectrometry, *Analytical chemistry 81*, 5966.

112. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange, *Proteins: structure, function, and bioinformatics 17*, 75-86.

113. Marcsisin, S. R., and Engen, J. R. (2010) Hydrogen exchange mass spectrometry: what is it and what can it tell us?, *Analytical and bioanalytical chemistry 397*, 967-972.

114. Marcsisin, S. R., and Engen, J. R. (2010) Molecular insight into the conformational dynamics of the Elongin BC complex and its interaction with HIV-1 Vif, *Journal of molecular biology 402*, 892-904.

115. Chetty, P. S., Mayne, L., Lund-Katz, S., Stranz, D., Englander, S. W., and Phillips, M. C. (2009) Helical structure and stability in human apolipoprotein A-I by hydrogen exchange and mass spectrometry, *Proceedings of the national academy of sciences of the United States of America 106*, 19005-19010.

116. Del Mar, C., Greenbaum, E. A., Mayne, L., Englander, S. W., and Woods, V. L., Jr. (2005) Structure and properties of alpha-synuclein and other amyloids determined at the amino acid level, *Proceedings of the national academy of sciences of the United States of America 102*, 15477-15482.

117. Zhu, S., Shala, A., Bezginov, A., Sljoka, A., Audette, G., and Wilson, D. J. (2015) Hyperphosphorylation of intrinsically disordered tau protein induces an amyloidogenic shift in its conformational ensemble, *PloS one 10*, e0120416.

118. Hansen, J. C., Wexler, B. B., Rogers, D. J., Hite, K. C., Panchenko, T., Ajith, S., and Black, B. E. (2011) DNA binding restricts the intrinsic conformational flexibility of methyl CpG binding protein 2 (MeCP2), *The Journal of biological chemistry 286*, 18938-18948.

119. Englander, J. J., Calhoun, D. B., and Englander, S. W. (1979) Measurement and calibration of peptide group hydrogen-deuterium exchange by ultraviolet spectrophotometry, *Analytical biochemistry 92*, 517-524.

120. Gregory, R. B., Crabo, L., Percy, A. J., and Rosenberg, A. (1983) Water catalysis of peptide hydrogen isotope exchange, *Biochemistry 22*, 910-917.

121. Englander, S. W. (2006) Hydrogen exchange and mass spectrometry: A historical perspective, *Journal of the American Society for Mass Spectrometry 17*, 1481-1489.

122. Tokuriki, N., and Tawfik, D. S. (2009) Protein dynamism and evolvability, *Science 324*, 203-207.

123. Hvidt, A., and Nielsen, S. O. (1966) Hydrogen exchange in proteins, *Advances in protein chemistry 21*, 287-386.

124. Skinner, J. J., Lim, W. K., Bedard, S., Black, B. E., and Englander, S. W. (2012) Protein hydrogen exchange: testing current models, *Protein science : a publication of the Protein Society 21*, 987-995.

125. Skinner, J. J., Lim, W. K., Bedard, S., Black, B. E., and Englander, S. W. (2012) Protein dynamics viewed by hydrogen exchange, *Protein science : a publication of the Protein Society 21*, 996-1005.

126. Konermann, L., Tong, X., and Pan, Y. (2008) Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches, *Journal of mass spectrometry 43*, 1021-1036.

127. Krishna, M. M., Hoang, L., Lin, Y., and Englander, S. W. (2004) Hydrogen exchange methods to study protein folding, *Methods (San Diego, Calif.) 34*, 51-64.

128. Wales, T. E., and Engen, J. R. (2006) Hydrogen exchange mass spectrometry for the analysis of protein dynamics, *Mass spectrometry reviews 25*, 158-170.

129. Keppel, T. R., and Weis, D. D. (2013) Analysis of disordered proteins using a simple apparatus for millisecond quench-flow H/D exchange, *Analytical chemistry 85*, 5161-5168.

130. Rob, T., Liuni, P., Gill, P. K., Zhu, S., Balachandran, N., Berti, P. J., and Wilson, D. J. (2012) Measuring dynamics in weakly structured regions of proteins using microfluidics-enabled subsecond H/D exchange mass spectrometry, *Analytical chemistry 84*, 3771-3779.

131. Connelly, G. P., Bai, Y., Jeng, M. F., and Englander, S. W. (1993) Isotope effects in peptide group hydrogen exchange, *Proteins 17*, 87-92.

132. Wang, L., Pan, H., and Smith, D. L. (2002) Hydrogen exchange-mass spectrometry: optimization of digestion conditions, *Molecular & cellular proteomics : MCP 1*, 132-138.

133. Busby, S. A., Chalmers, M. J., and Griffin, P. R. (2007) Improving digestion efficiency under H/D exchange conditions with activated pepsinogen coupled columns, *International journal of mass spectrometry 259*, 130-139.

134. Mazon, H., Marcillat, O., Forest, E., and Vial, C. (2005) Denaturant sensitive regions in creatine kinase identified by hydrogen/deuterium exchange, *Rapid communications in mass spectrometry : RCM 19*, 1461-1468.

135. Nazabal, A., Maddelein, M.-L., Bonneu, M., Saupe, S. J., and Schmitter, J.-M. (2005) Probing the structure of the infectious amyloid form of the prion-forming domain of HET-s using high resolution hydrogen/deuterium exchange monitored by mass spectrometry, *The Journal of biological chemistry 280*, 13220.

136. Fitzgerald, M. C., and West, G. M. (2009) Painting proteins with covalent labels: what's in the picture?, *Journal of the American society for mass spectrometry 20*, 1193-1206.

137. Xu, G., and Chance, M. R. (2004) Radiolytic modification of acidic amino acid residues in peptides: probes for examining protein-protein interactions, *Analytical chemistry 76*, 1213-1221.

138. Xu, G., and Chance, M. R. (2007) Hydroxyl radical-mediated modification of proteins as probes for structural proteomics, *Chemical reviews 107*, 3514-3543.

139. Sharp, J. S., Becker, J. M., and Hettich, R. L. (2003) Protein surface mapping by chemical oxidation: structural analysis by mass spectrometry, *Analytical biochemistry 313*, 216-225.

140. Guan, J. Q., Vorobiev, S., Almo, S. C., and Chance, M. R. (2002) Mapping the G-actin binding surface of cofilin using synchrotron protein footprinting, *Biochemistry 41*, 5765-5775.

141. Sharp, J. S., Guo, J. T., Uchiki, T., Xu, Y., Dealwis, C., and Hettich, R. L. (2005) Photochemical surface mapping of C14S-Sml1p for constrained computational modeling of protein structure, *Analytical biochemistry 340*, 201-212.

142. Hambly, D. M., and Gross, M. L. (2005) Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale, *Journal of the American society for mass spectrometry 16*, 2057-2063.

143. Hambly, D., and Gross, M. (2007) Laser flash photochemical oxidation to locate heme binding and conformational changes in myoglobin, *International journal of mass spectrometry 259*, 124-129.

144. Gau, B. C., Sharp, J. S., Rempel, D. L., and Gross, M. L. (2009) Fast photochemical oxidation of protein footprints faster than protein unfolding, *Analytical chemistry 81*, 6563-6571.

145. Kaur, P., Kiselar, J. G., and Chance, M. R. (2009) Integrated algorithms for high-throughput examination of covalently labeled biomolecules by structural mass spectrometry, *Analytical chemistry 81*, 8141-8149.

Chapter Two

Mechanism of calmodulin-induced activation of calcineurin revealed using hydrogen exchange mass spectrometry

2.1 Introduction

Calcineurin (CaN) is a heterodimeric serine/threonine phosphatase activated by calcium and calmodulin (CaM) binding.[1-4] CaN plays essential roles in various cellular and biological processes such as T-cell activation, muscle contraction, neuronal signaling, and cardiac growth.[3] At high calcium concentration in the cell cytoplasm, calcium ions activate calmodulin, resulting in calmodulin binding to CaN and causing to CaN activation.[5] One of the common substrates of CaN is the nuclear factor of activated T cells (NFAT) family of transcription factors.[6, 7] CaN binds directly to NFAT, leading to dephosphorylation of NFAT. The dephosphorelation is followed by translocation of NFAT into the cell nucleus, resulting in DNA transcription.[7, 8] CaN dysregulation has been linked to a number of major diseases such as Alzheimer's disease and Down's syndrome.[9, 10] Due to its role in many biological processes, CaN has been the target for two immunosuppressant drugs, tacrolimus (FK506) and cyclosporine A (CsA).[11-14] FK506 and CsA bind tightly to the abundant intracellular proteins FKBP12 and cyclophilin A, respectively, resulting in activation of these proteins. These drug-protein complexes then inhibit CaN activation in the cell by directly binding to CaN.

CaN is a heterodimer composed of two subunits: calcineurin A chain (CaN A) with molecular mass of 59 kDa and calcineurin B chain (CaN B) with a mass of 19 kDa.[15] The ribbon diagram of CaN structure from PDB entry 1AUI[16] is shown in Figure 2.1b. Human CaN A has three isoforms (α, β, and γ) and CaN B has two isoforms (1 and 2).[17] α and β CaN A are the two main isoforms that bind to a common CaN B1 isoform. α CaN is the most dominant form in neurons, the βCaN isoform is broadly distributed, and the γ isoform binds to a testis-specific CaN B2 isoform. Based on biochemical and genetic studies, four distinct functional domains of CaN A have been identified.[18, 19] Figure 2.1a shows the domain organization of the CaN A. The

domains are a catalytic domain, a CaN B-binding domain (CaNBBD), a calmodulin-binding domain (CaMBD), and an auto-inhibitory domain (AID). Together, the calmodulin-binding domain and the AID comprise the regulatory domain (RD) of the CaN A chain. Except for a short ordered helix of the AID, all parts of the regulatory domain are absent in the electron density map of αCaN A (PDB: 1AUI).[16] Figure 2.1b shows the entire CaN molecule including a model of the disordered regulatory domain. Within the CaN A catalytic domain there are a zinc and an iron atom that are part of the catalytic active site. CaN B, which has 35% sequence identity with calmodulin, consists of two lobes with two calcium ions bound by EF-hand motifs in each lobe.

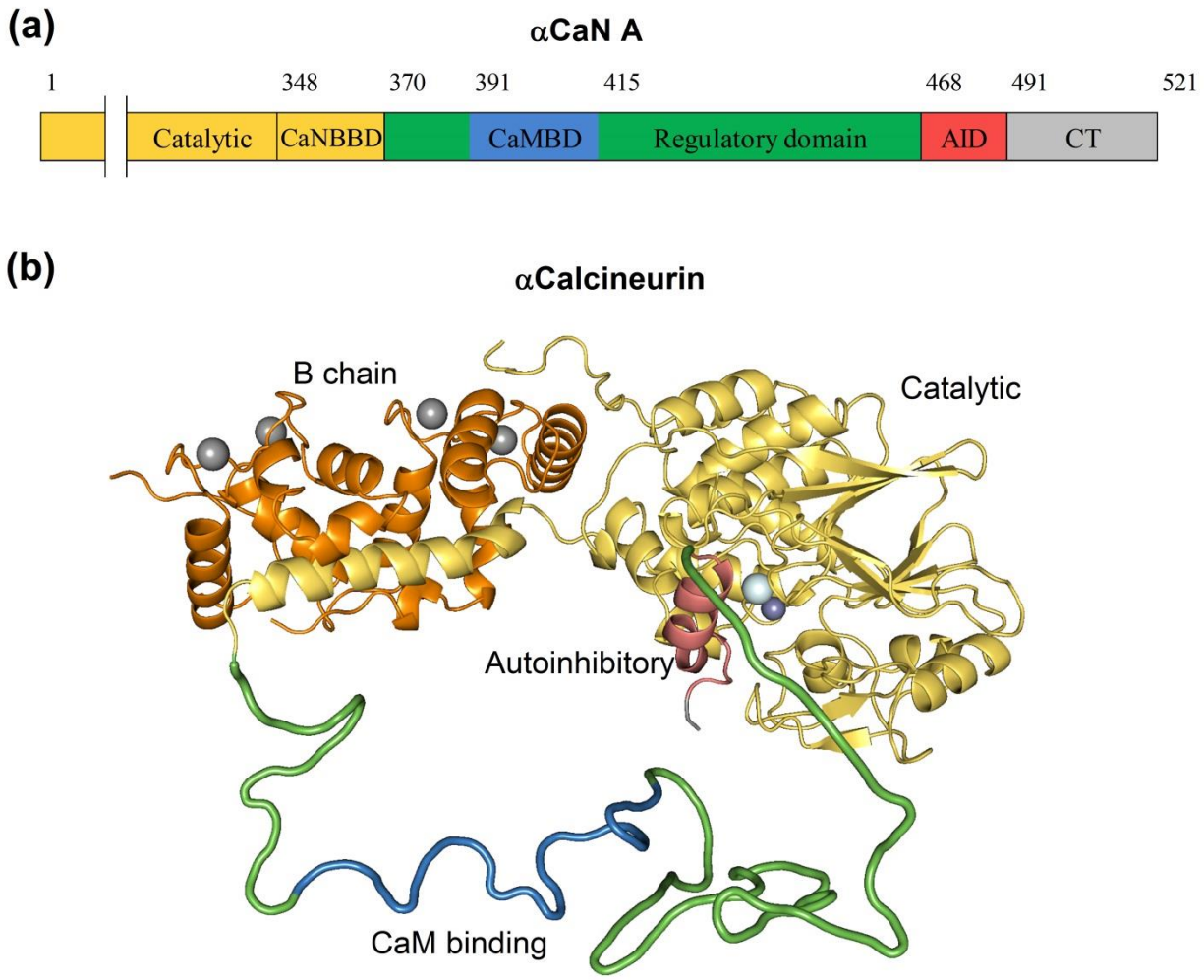


Figure 2.1. (a) Domain organization of the αCaN A chain and (b) structure of human α-calcineurin (CaN). The depiction of CaN structure is modified from PDB entry 1AUI.[16] The CaN A and CaN B are colored in yellow and orange, respectively. The autoinhibitory domain is shown in pink. The disordered regulatory domain is shown in green. The calmodulin binding domain is shown in blue. The disordered regulatory domain, which is absent from the crystal structure, was modeled with PyMOL (Schrödinger LLC, Portland, OR, USA).

Depending on the calcium concentrations in the cell, CaN undergoes sequential conformation changes to become fully activated.[15] Several models of CaN activation have been proposed, and a widely accepted model of activation is explained here.[5, 15] At low calcium concentrations, CaN exists in an inactive state in which only the high-affinity calcium binding site of CaN B is occupied with calcium. In addition, the AID and the calmodulin-binding domain

remain associated with the active site and the CaN B-binding domain, respectively.[5] At high calcium concentration in the absence of CaM, calcium binds the low-affinity sites on CaN B, resulting in the dissociation of a major part of CaN regulatory domain including the calmodulin-binding domain causing CaN to become partially active. At high calcium concentration and in the presence of CaM, CaM binds four calcium ions and then binds to the calmodulin-binding domain of CaN. This binding causes an ordering in the regulatory domain that removes the AID from the CaN active site, turning CaN into a fully active state.[20, 21] Many techniques have been used to characterize the CaN activation mechanism including X-ray crystallography,[16] fluorescence,[22] limited tryptic digestion,[23] size-exclusion chromatography,[24] small-angle X-ray scattering,[25] and hydrogen exchange.[21, 26] Some of the studies are based on low-resolution techniques to characterize truncated forms of CaN where the regulatory domain sequence is absent from the structure. Some other studies are only based on high-resolution techniques to describe the regulatory domain to probe the CaN structural changes. None of the high-resolution techniques have been used to describe the CaN structural change in solution using full length CaN. In addition, working with the truncated forms of proteins sometimes is inadequate because they may not reliably represent how full-length forms undergo conformational change. Thus, more structural information about conformational changes in solution is required to fully understand the mechanism of CaN activation.

Our previous work on the regulatory domain-autoinhibitory domain-C-terminal domain (RD-AID-CT) construct isolated from full length CaN showed that the regulatory domain and AID are completely disordered at high calcium concentration in the absence of CaM.[21] At high calcium concentration in the presence of CaM, two regions of the regulatory domain become structured. The calmodulin-binding domain becomes tightly structured. The region located to the

C-terminal side of calmodulin-binding domain becomes weakly structured. The data suggested that the structuring of the two regions of the regulatory domain causes displacement of the AID from the active site, but CaM has no effect on the AID, which remains unstructured. Therefore the AID must become structured only when it binds to the active site. To better understand the mechanism of activation, studies on full-length CaN in solution are required. Recently, there has also been considerable debate about the stoichiometry of CaN-CaM complex.[24, 25, 27] The questions of whether there is a second weak CaM binding site in the regulatory domain of CaN and what is the binding ratio of the complex in solution need to be addressed.

Hydrogen exchange (HX) mass spectrometry can provide vital information about protein structural changes in solution at peptide level resolution. This technique is based on labeling the backbone amide hydrogens of a protein with deuterium. The rate of exchange is a function of the structure and dynamics of the protein because the exchange is highly dependent on hydrogen bonding and solvent accessibility.[28, 29] Highly flexible regions will exchange more rapidly, while rigid regions exchange more slowly.[28] Protection in HX experiments, which is a measure of the slowing of exchange caused by structure, can reveal information about protein structures. In this work, we used full-length CaN to understand the mechanism of CaN activation upon calmodulin binding in the presence of Ca^{2+}. The results with full-length CaN demonstrate that the calmodulin-binding domain of the regulatory domain is completely disordered in the absence of CaM, while it folds tightly upon binding to calmodulin in the presence of excess of Ca^{2+}. This result confirms our previous work on isolated RD-AID-CT.[21] However, using full-length CaN, we did not observe any weak protection in the region adjacent to the C-terminal side of the calmodulin-binding domain. Our data also reveal that a region in the regulatory domain that is immediately adjacent to the N-terminal side of the AID becomes weakly protected. Additionally,

we have observed CaM-induced increase in flexibility in segments located near the CaN active site. The observed HX kinetics also show that some segments of the B chain become protected upon calmodulin binding. The data suggest that calmodulin, which binds to the regulatory domain in the CaN A, may interact with the CaN B and the N-terminal side of the AID. Finally, and most surprisingly, we found no changes in the structuring of the CaN AID upon CaM binding. We present results from all regions of the CaN A and B chains to describe the mechanism of CaN activation. We also propose a new model of CaN activation upon CaM binding. In a separate HX experiment, we also explore several possible binding ratios of CaN to CaM using HX mass spectrometry on the isolated RD-AID construct from CaN. We confirmed that the stoichiometry of CaN:CaM binding in solution is 1:1.

2.2 Materials and methods

2.2.1 Materials

Calcium chloride, porcine pepsin, deuterium oxide (D_2O) (99.9% D), DL-dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), and sodium chloride were purchased from Sigma (St. Louis, MO). LC-MS grade formic acid (99+ %) and water used in chromatography were purchased from Fisher Scientific (Hanover Park, IL). Pre-mixed 0.1% formic acid in acetonitrile (v/v) was purchased from Honeywell VWR (Arlington Heights, IL). Sodium hydroxide and hydrogen chloride were both from Fisher Scientific.

2.2.2 Protein expression and purification

Full-length human α calcineurin (CaN), a truncated regulatory domain-autoinhibitory domain-C-terminal construct (RD-AID-CT), and human calmodulin (CaM) were gifts from Prof. Treavor Creamer (University of Kentucky). All the proteins were expressed and purified as described previously.[21] Briefly, CaN was expressed from a pETagHisCN expression vector

containing CaN A (with N-terminal His_6 tag) and B1 genes in *Escherichia coli* BL21 (DE3) CodonPlus RIL cells. The RD-AID-CT construct was expressed from a pET303/CT-His expression vector containing a gene for the human sequence RD-AID-CT in *Escherichia coli* BL21 (DE3). Both proteins were purified by using a Ni-NTA column followed by a CaM-Sepharose[(30)] column. Calmodulin was expressed from a pETCaMI expression vector in *Escherichia coli* BL21 (DE3). After expression, CaM was purified on a 2-triflouromethyl-10-aminopropyl phenothiazine-Sepharose column.[(30)] Protein concentrations of CaN (13.9 µM), RD-AID-CT (20.3 µM), and CaM (100 µM) were determined using a BCA assay[(31)] standardized with BSA (Thermo Fisher Scientific). The identity of each protein was confirmed by intact mass spectrometry measurements.

2.2.3 Reduction of calcineurin B-dimerization and enzyme kinetics

In the mass spectrum of CaN, we observed both monomeric B at 19168.6 Da (theoretical mass of 19168.7 Da) and a dimer at 38335.6 Da consistent with a disulfide-bonded dimer (theoretical mass of 38335.4 Da)(see deconvoluted spectra in Figure 2.S1). Since CaN B was not supposed to have disulfide bonds because it has two cysteines in the form of free thiols [(16)], we used DL-dithiothreitol (DTT) to reduce CaN B dimerization. This step was necessary to make the protein ready for hydrogen exchange analysis. The intact CaN (0.4 µM) was prepared in two buffers, one without DTT as a control (20 mM HEPES, 100 mM NaCl, 10 mM $CaCl_2$, pH 7.5) and the other with DTT (5 mM DTT, 20 mM HEPES, 100 mM NaCl, 10 mM $CaCl_2$, pH 7.5). CaN was incubated in DTT for 15 min, 40 min, and 2 h at 4 °C in a thermostated block (IC20 Ecotherm Heating/Chilling Plate, Torrey Pines Scientific, La Jolla, CA). The samples were then loaded onto an HPLC system (Agilent 1200 series) combined with time-of-flight mass spectrometer (Agilent 6220, Santa Clara, CA, USA). The mobile phases were 0.1% formic acid

(A) and 90% acetonitrile/10% water/ 0.1% formic acid (B). The CaN A, CaN B, and dimer of CaN B were trapped and desalted on a C_4 trap (Jupiter C4 self-packed 1 mm × 10 mm cartridge, Phenomenex, Torrance CA) for 3 min at 200 µL/min A followed by separation on a C_4 column (Jupiter 4u Proteo 90 A°, 4 micron, 1 × 50 mm) using a gradient 15-60% B over 20 min at 50 µL/min. To monitor the reduction of CaN B dimer and increasing the intensity of monomer CaN B, we compared the area under the peak of extracted ion chromatogram (EIC) of 913.78 m/z with 892.50 m/z which correspond to monomer CaN B $[M+21H]^{21+}$ and dimer CaN B $[2M-2H+43H]^{43+}$, respectively (see mass spectra in Figure 2.S1). The data summarized in Table 2.1 show that the amount of CaN B dimer decreases by increasing the protein incubation time in DTT. Since we were using 5 mM DTT as a reducing agent, we wanted to compare the CaN enzymatic activity under non-reducing with reducing conditions. This comparison was necessary to confirm that DTT did not damage the enzymatic activity of CaN. A CaN activity assay was then performed by our collaborators in the Trevor Creamer lab (University of Kentucky). The activity was assayed using p-nitrophenyl phosphate (pNPP) as the substrate. The assay was performed with CaN diluted in the assay buffer (50 mM Tris pH 7.5, 100 mM NaCl, 6 mM $MgCl_2$, 0.5 mM DTT, 0.025-40% pNPP, 0.5 mM $CaCl_2$) and with CaN in the same buffer but with a 40 min preincubation in 5 mM DTT. The reactions were performed at 37 °C. The absorbance of the reactions at 405 nm was measured immediately and after 2, 5, 10, 20, 30, 40 minutes. According to the assay, the activity of CaN in the absence of DTT is comparable to CaN in the presence of DTT in the buffer even after 40 min of the enzyme-substrate reaction (see Figure 2.2). Since we preincubated CaN in DTT for 40 minutes for the activity assay and showed no difference in activity without the added DTT, we proceeded to incubate CaN in 5 mM DTT for 40 min before hydrogen exchange experiments.

Table 2.1. Reduction of calcineurin B-dimerization using dithiothretol (DTT). The intact CaN sample was prepared in two buffers, one without DTT as a control (20 mM HEPES, 100 mM NaCl, 10 mM $CaCl_2$, pH 7.5) and the other with 5 mM DTT in the same buffer. CaN was incubated in DTT for 15 min, 40 min, and 2 h at 4 °C to determine the effective condition to reduce CaN B dimer. The EIC of 913.785 and 892.508 m/z correspond to monomer CaN B $[M+21H]^{21+}$ and dimer CaN B $[2M-2H+43H]^{43+}$, respectively. The area under peak (AUP) for each EIC at each condition was measured. At each condition the area under peak was converted to the percentage value, $\%AUP_{momomer}=[AUP_{momomer}/(AUP_{momomer}+ AUP_{dimer})] \times 100$.

condition	Area under peak	
	monomer (913.785 m/z)	dimer (892.508 m/z)
CN- DTT	78.3	21.6
CN+15 min DTT	89.9	10.0
CN+40 min DTT	92.4	7.5
CN+2 h DTT	96.5	3.4

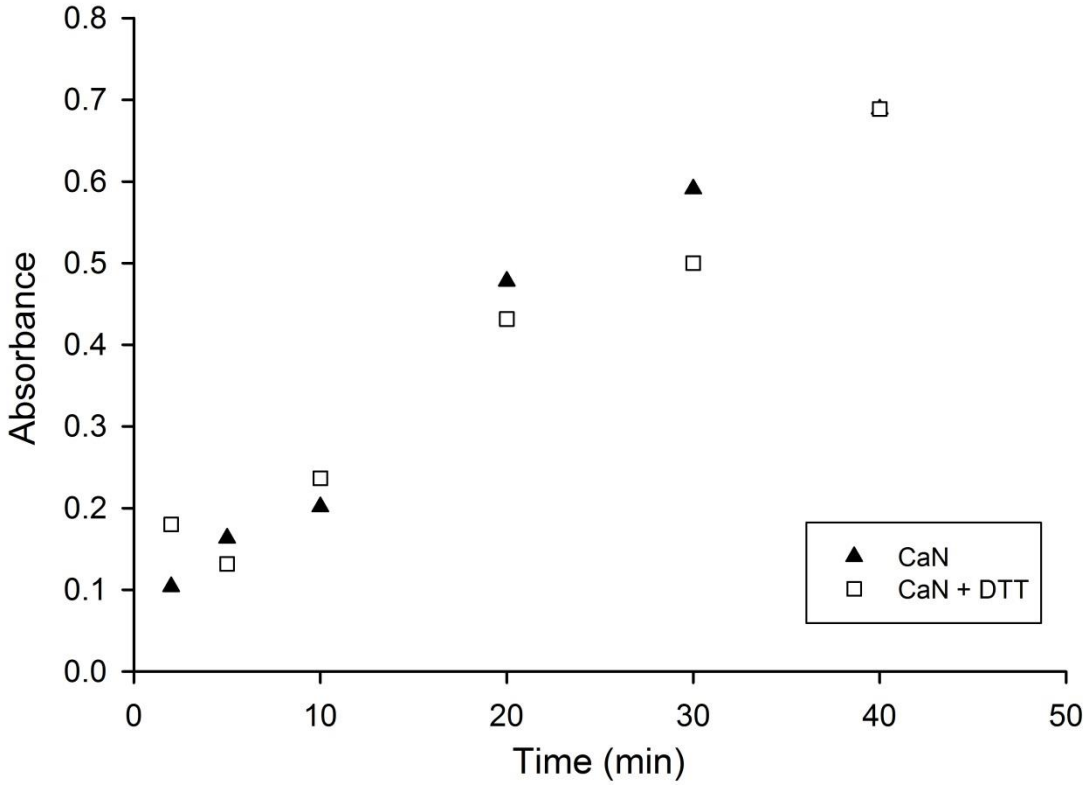


Figure 2.2. Activity assay of calcineurin in the absence and presence of 5 mM DTT. The assay performed using para-nitrophenyl phosphate (*p*NPP) in the assay buffer (50 mM Tris pH 7.5, 100 mM NaCl, 6 mM $MgCl_2$, 0.5 mM DTT, 0-200 mM pNPP, 0.5 mM $CaCl_2$). The activity of CaN without DTT and CaN preincubated in 5 mM DTT for 40 min was assayed. The absorbance of nitrophenyl at 405 nm was then measured. The vertical axis is the absorbance and the horizontal axis is the length of the activity assay.

2.2.4 Hydrogen exchange mass spectrometry

2.2.4.1 Calcineurin peptide identification

CaN peptic peptides were assigned using a combination of a time-of-flight mass spectrometer (Agilent 6220) and MS/MS enabled in a linear ion trap (Thermo LTQ-XL). In the first step, CaN sample (20 pmol) was loaded onto an HPLC system (1200 series) combined with a TOF mass spectrometer. The protein samples were digested online at 200 uL/min with mobile phase A using an immobilized pepsin column[32, 33] (self-packed in a 2.1 mm × 100 mm column). The resulting peptides were captured and desalted on a C12 trap column over 4 min. The peptides were then separated using the following gradient: 5% B (0 min), 15% B (1 min), 45% B (9 min), 60% B (10 min), 95% B (11 min), 95% B (13 min), and 5% B (14 min) at 50 µL/min B. The peptides were identified on the basis of accurate mass (10 ppm mass tolerance) measurements using 4-GHz high-resolution mode of the TOF mass spectrometer. The peptides that could not be unambiguously assigned on the basis of accurate mass were assigned in a separate MS/MS experiments. The CaN peptic peptides were first prepared by offline digestion using an immobilized pepsin column at 200 µL/min at 4 °C using a syringe pump. The peptides were then collected and flash frozen with liquid nitrogen. Identity confirmation experiments on the basis of the collision-induced dissociation fragmentation were performed on a linear ion trap mass spectrometer combined HPLC (Shimadzu, Columbia, MD). The peptides were separated by HPLC on a C18 column (ZORBAX SB300-C18, 2.1 × 100 mm, Agilent) with a 110 min 0-60% B gradient (A: 99% H_2O, 1% ACN, 0.1% formic acid; B: 99% CAN, 1% H_2O, 0.1% formic acid; 200 µL/min flow rate). Mass spectra were analyzed using the Proteome Discover 1.3 software (Thermo Scientific). Only MS/MS spectra with XCorr score of ≥2 were used to validate peptide assignments. A list of all assigned peptides from both set of experiments was made for the

hydrogen exchange analysis. A peptic peptide map, shown in Figure 2.3, covering 95% of each of the chain of CaN, CaN A and CaN B, was constructed from a total of 134 peptides that were reproducibly obtained by pepsin digestion of CaN.

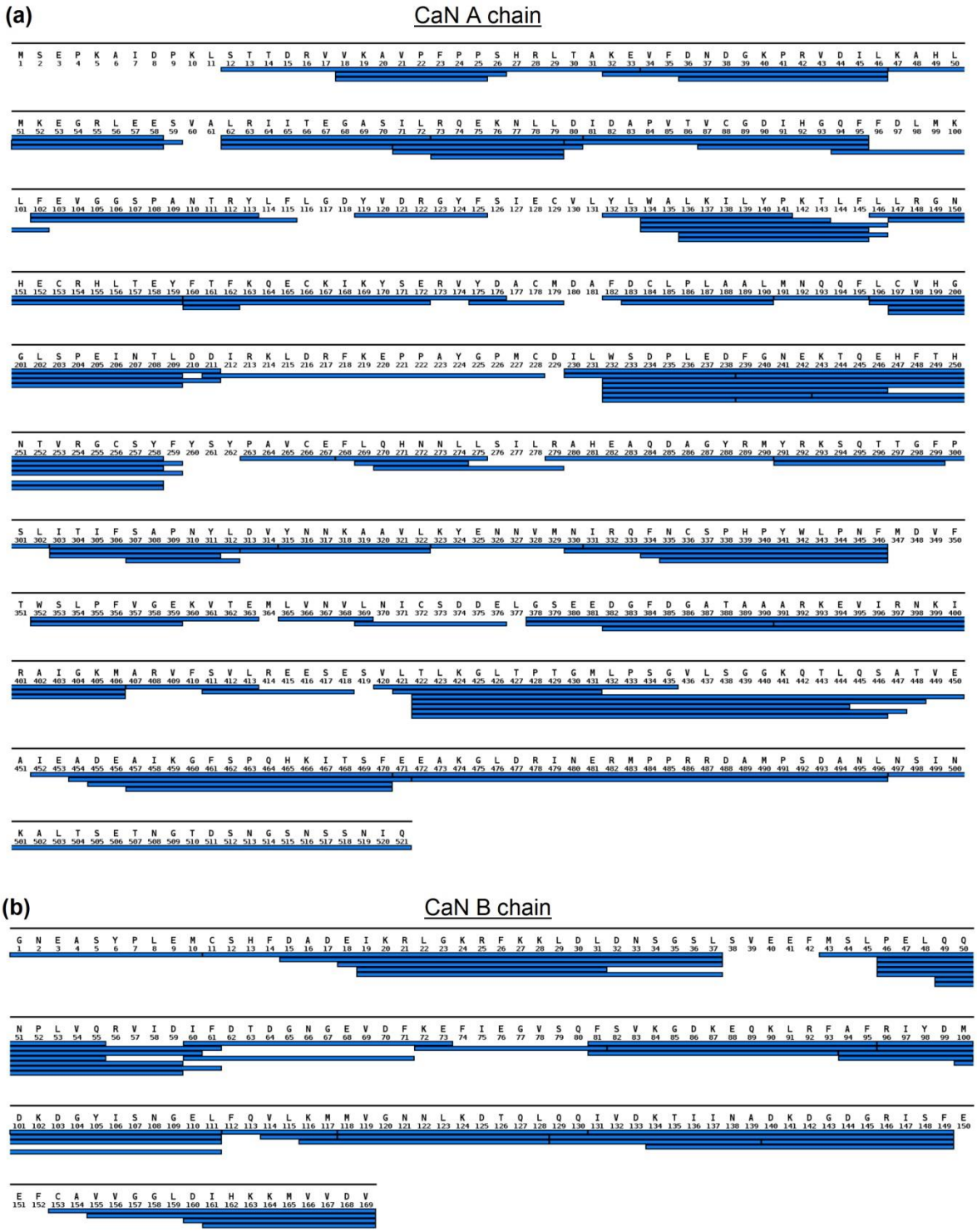


Figure 2.3. Peptic peptide map of the (a) calcineurin A chain and (b) calcineurin B chain. The coverage map composed of 134 peptides covering 95% of each of the CaN A and CaN B primary sequence. These peptides were reproducibly obtained by pepsin digestion of CaN. The coverage map was constructed using MS Tools.[34]

2.2.4.2 Calcineurin (RD-AID-CT)-calmodulin binding ratio

To determine the calcineurin (RD-AID-CT):calmodulin binding ratio, we performed HX using three different molar ratios of CaM 1:1, 1:2, and 1:3 and RD-AID-CT without CaM. The samples (12 µM RD-AID-CT (control) or in mixture with CaM at 12, 24, 36 µM) were prepared in H_2O buffer (20 mM HEPES, 100 mM NaCl, 10 mM $CaCl_2$, pH 7.5) and incubated at 4 °C for 1 h before deuterium labeling. We used an H/DX PAL (LEAP Technologies, Carrboro, NC) robot system for sample labeling and quenching. We used one labeling time point (12 s) to monitor the protection from exchange according to different binding ratios. 3 µL of the samples were diluted 19-fold with D_2O buffer (20 mM HEPES, 100 mM NaCl, 10 mM $CaCl_2$, pD 7.5) and quenched with 1% formic acid by 1:1 dilution at 1 °C. For fully deuterated control, the RD-AID-CT sample was incubated in D_2O buffer for 24 hours before quenching the sample. 90 µL of the quenched solution was injected into the sample loop of a refrigerated compartment which includes an immobilized pepsin column, a peptide desalting trap (Porshell 120 EC-C18, 2.1 × 5 mm, 2.7 µm particles, Agilent Technologies), and a C18 (Zorbax 300SB-C18, 1.8 µm particles, 2.1 × 50 mm, Agilent Technologies) column. The compartment was held at 1 °C for this experiment. All LC-MS analysis of the samples were carried out on a quadrupole time-of-flight mass spectrometer (Agilent 6530, Santa Clara, CA, USA) combined with a multipump LC system (Agilent 1260).

2.2.4.3 Calcineurin-calmodulin hydrogen exchange

We used an H/DX PAL (LEAP Technologies) robot system for sample handling for HX experiments. Calcineurin (CaN) (12 µM) containing 5 mM DTT was incubated with calmodulin (CaM) (12 µM) or an equal volume of H_2O buffer (20 mM HEPES, 100 mM NaCl, 10 mM $CaCl_2$, pH 7.5) for 40 minutes at 4 °C. The samples were then transferred to the robotic system

for labeling. 3 µL of the samples were diluted 19-fold with D_2O buffer (20 mM HEPES, 100 mM NaCl, 10 mM $CaCl_2$, pD 7.5) for between 13 s and 24 h at 25 °C. For a non-deuterated control, the samples were prepared similarly and diluted with H_2O buffer instead of D_2O buffer. The samples were quenched to a pH of 2.6 by addition of 1% formic acid at 1:1 dilution ratio. After quenching, the samples were loaded into the sample loop of the refrigerated compartment. To obtain more efficient pepsin digestion, the HPLC lines were extended to set the immobilized pepsin column outside the compartment in order to digest proteins at room temperature (23±1°C). The desalting trap and the reversed phase column remained inside the compartment where the temperature was maintained at 1 °C. After injection, the samples were digested online at 200 µL/min with 0.1% formic acid. The resulting peptides were captured and desalted on a reversed-phase trap (1 mm × 10 mm self-packed with Jupiter Proteo C12 media obtained from Phenomenex) over 5 min. The peptides were eluted from the trap and separated using a C18 column (Zorbax 300SB-C18, 1 mm × 50 mm, 3.5 µm particles, Agilent Technologies). For gradient elution the mobile phases were 0.1% formic acid (A) and 90% ACN:10% water:0.1% formic acid (B). The gradient was as follows: 5% B (0 min), 15% B (1 min), 45% B (9 min), 60% B (10 min), 95% B (11 min), 95% B (13 min), and 5% B (14 min) at 50 µL/min B. The LC-MS analysis was carried out on a time-of-flight mass spectrometer (Agilent 6220) combined with a HPLC (Agilent 1200). The mass spectra were collected in positive ESI mode with a capillary voltage of 4000 V, fragmentor of 150 V, and temperature of 325 °C.

2.2.4.4 Data Analysis

Mass spectra of the CaN peptides were analyzed using Agilent MassHunter Qualitative Analysis software (version B.06.00, Agilent). HDExaminer (version 1.0, Sierra Analytics, Modesto, CA) was used for initial deuterium calculation and data analysis. We processed three

replicate sets of HX data of CaN A and CaN B in the presence and absence of CaM to determine standard deviation values in the HX analysis. We used an R script, written in house, for deuterium uptake data averaging and plotting for a total of 134 peptides for CaN A and B. The standard deviation (SD) at each labeling time point and for each state of CaN was determined. We followed a method described by Houde et al. to determine statistical values for replicate differential HX data.[(35)] A histogram of the statistical analysis of the data is shown in Figure 2.4. A statistical analysis of our triplicate measurements established that the 98th percentile of the standard deviations ($SD_{98\%}$) was ±0.4 Da. For differential HX measurements, i.e., $\sum \Delta m = \sum m_{CaN+CaM} - \sum m_{CaN}$, the 98% confidence limit was then propagated to account for all the time points together (9 time points for each state multiplied by 2) and the overall value, $\sqrt[2]{18 \times SD_{98\%}^2}$, was 1.7 Da. Thus, if the summed deuterium difference is greater than 1.7 Da, this difference can be considered statistically significant at the 98% confidence level. All significant differences in deuterium content between two states of CaN were mapped onto the crystal structure of CaN.

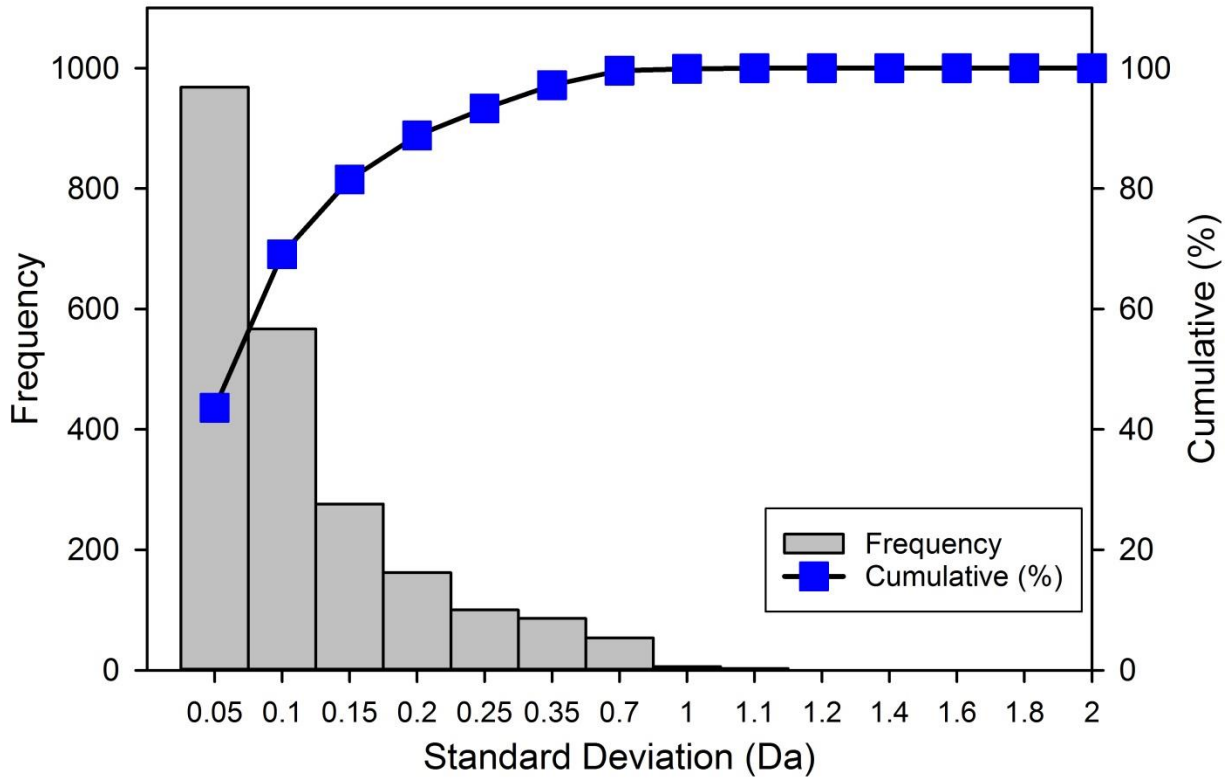


Figure 2.4. Reproducibility of hydrogen exchange mass spectrometry data represented by the distribution of standard deviations for the mass differences across all time points and states of the calcineurin peptides from the triplicate experiments (N = 2222). The 98th percentile for standard deviations was 0.4 Da.

2.3 Results

2.3.1 Stoichiometry of CaM:CaN RD-AID-CT is 1:1

Hydrogen exchange (HX) mass spectrometry was conducted to determine the stoichiometry of CaM:CaN complex. The complex has very tight binding with a K_d value in the picomolar range.[24, 36, 37] In order to determine the binding ratio in solution, we monitored protection from exchange by CaN RD-AID-CT in the absence (control) and presence of CaM at three different binding ratios (CaN RD-AID-CT: CaM): 1:1, 1:2, and 1:3. We used one time point (12 s) to monitor the changes in HX for each peptide at each condition. The peptide deuterium content at each condition was converted to deuteration percentage to obtain an HX-

MS heat map grouped by representative peptides from various regions of RD-AID-CT (see Figure 2.5). The deuterium uptake data were corrected for back exchange using a fully deuterated control. When regulatory domain was in the free state, we observed full deuteration in the CaM binding site for the peptide A 382-406 at 12 s. The deuteration percentage decreased to 45% when regulatory domain bound to CaM at a 1:1 ratio for the peptide at the same time point. This result indicates that this region becomes significantly protected when CaM binds. We observed the same percentage of deuteration in the presence of CaM at two or three-fold molar ratios. The data indicate that CaM binds the calmodulin-binding domain of regulatory domain at a 1:1 binding ratio. In addition, we monitored the differences in deuteration percentages in the other regions of regulatory domain. We found that there are no differences in the exchange in the presence of CaM at different binding ratios. All the data indicate that the binding ratio of CaM:CaN RD-AID-CT is 1:1. These results are consistent with our previous measurements of HX in the RD-AID-CT construct carried out at 1:2.[(21)] Thus, subsequent experiments with full-length CaN in complex with CaM were prepared at 1:1.

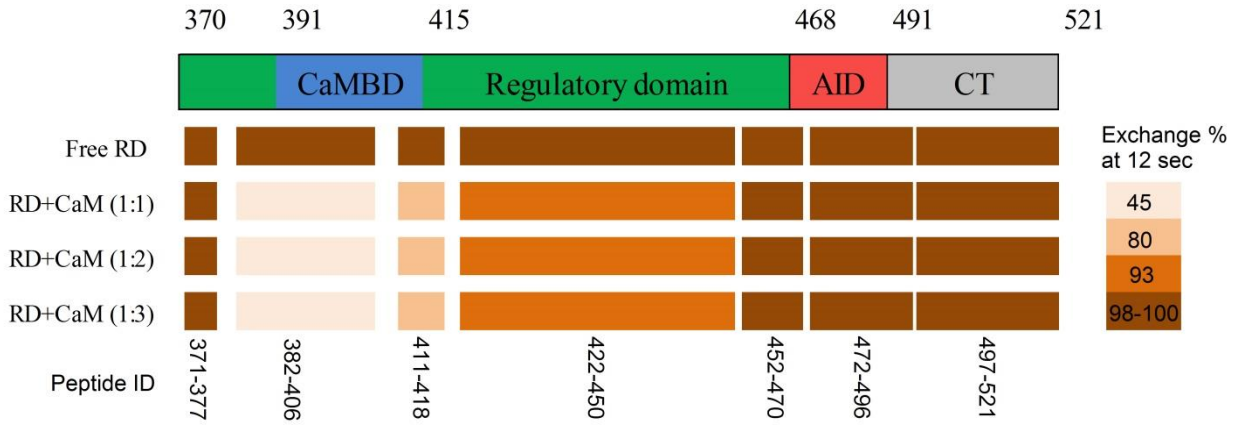


Figure 2.5. Deuterium percentage heat map for the isolated calcineurin RD in the absence and presence of a one-fold, two-fold, and three-fold molar excess of calmodulin at 12 s labeling time point. The graph constructed from non-overlapping peptides representative from different regions of RD to monitor the changes in deuteration. The domain organization of the isolated RD is shown at top for reference. The representative peptide identity is shown at the end of each bar. The deuteration percentage in the presence calmodulin is compared to the deuteration percentage of free RD state (control).

2.3.2 Effects of CaM binding to full length CaN

Hydrogen exchange mass spectrometry (HX-MS) was used to assess the effects of CaM binding on full length CaN conformation. Two sets of HX experiments on full-length CaN in the absence and presence of CaM at 1:1 binding ratio were run (see Materials and Methods section). The protein samples were incubated in deuterium-containing solutions for various time points followed by quenching the hydrogen-deuterium exchange reaction and then passed through a pepsin column. The deuterium content in the resulting peptides was measured by mass spectrometry (see Materials and Methods section). Data from representative peptides from various region of CaN A and B chains are shown in Figure 2.6. The deuterium uptake curves for all peptides of CaN in the absence and presence of CaM are presented in Figure 2.S2. The shapes of deuterium uptake profiles can provide information about flexibility in different regions of CaN. The differences in the rates of exchange between two states of the proteins can provide

information about CaN conformational change when CaM binds. We observed differences in the deuterium uptake between two states of CaN in several peptides from various regions of CaN. One other way to present the effects of CaM binding on CaN is differential deuterium uptake. The differential deuterium value of each peptide at each HX time point (13 s, 20 s, 40 s, 3 min, 12 min, 48 min, 2 h, 8 h, 24 h) was calculated by deducting the deuterium value of free CaN from that of bound CaN, as shown in Figure 2.7 for both chains. The peptide number on the *X*-axis is ordered from N-terminal to C-terminal region of CaN. The *Y*-axis is the summed difference for all HX time points for each peptides, $\sum \Delta m = \sum m_{\mathrm{CaN+CaM}} - \sum m_{\mathrm{CaN}}$. A positive value for $\sum \Delta m$ indicates that the CaM binding caused faster exchange which means there is a decrease in protection (more flexible), while a negative value indicates slower exchange which means there is an increase in protection (less flexible). To set the criteria for determining statistically significant differences between two protein states, the 98% confidence interval (± 1.7 Da) was determined and plotted as dotted lines in the differential deuterium uptake (Figure 2.7) (see Materials and Methods section for the method of determination). The local differences are mapped onto the crystal structure of CaN as shown in Figure 2.8. The CaM-induced increases in local protection are colored in yellow, decreases in protection in green, and no changes in gray. The absence of HX data is shown in white.

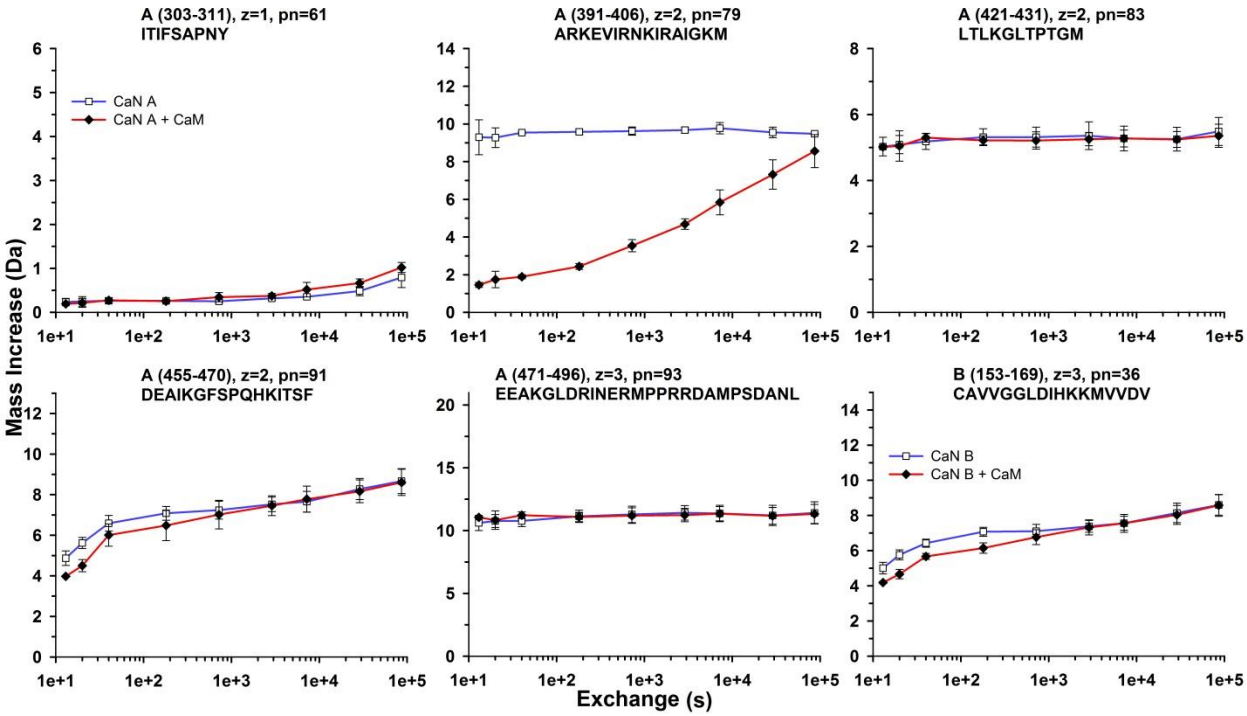


Figure 2.6. Deuterium uptake curves for six representative peptides from different regions of calcineurin A and B chains. The peptide identity, charge state, peptide number (pn) corresponding to pn in Figure 2.7a for A chain and 2.7b for B chain, and amino acid sequence are provided on each plot. The open squares represent uptake by the free calcineurin (CaN) state and the black filled diamonds represent uptake by CaN bound to calmodulin state. The error bars denote standard deviation from triplicate measurements. The horizontal axis in each plot is the exchange time in seconds plotted on logarithmic scale and the upper limit of the vertical axis is the maximum theoretical number of exchangeable amides in each peptide at 100% labeling.

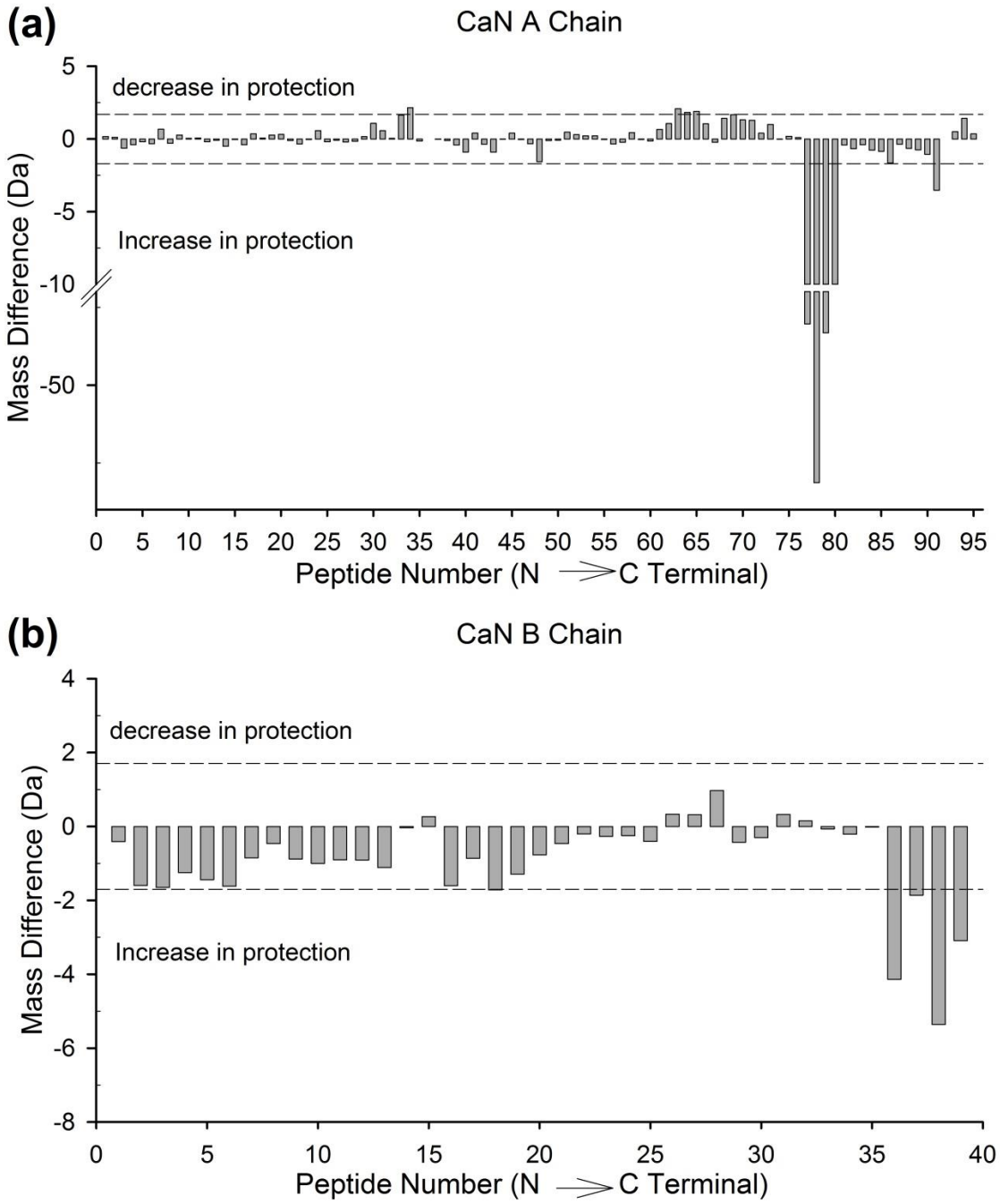


Figure 2.7. Mass difference plots of calcineurin (CaN) (a) A chain and (b) B chain in the presence of calmodulin compared with the absence of calmodulin as measured by HX-MS. Deuterium uptake differences for the two states of CaN were calculated for all exchange time points and summed for all labeling time points, $\sum \Delta m = \sum m_{\mathrm{CaN+CaM}} - \sum m_{\mathrm{CaN}}$. The summed differences are plotted on the vertical axes. The horizontal axes denote the peptides numbered from the N-to-C termini of the CaN A and CaN B. The dashed line at ±1.7 Da indicates the 98% confidence limit for significant mass differences after propagating the standard deviation to account for all the time points together. The positive and negative bars represent peptides that show decrease and increase in protection, respectively, in the presence of CaM.

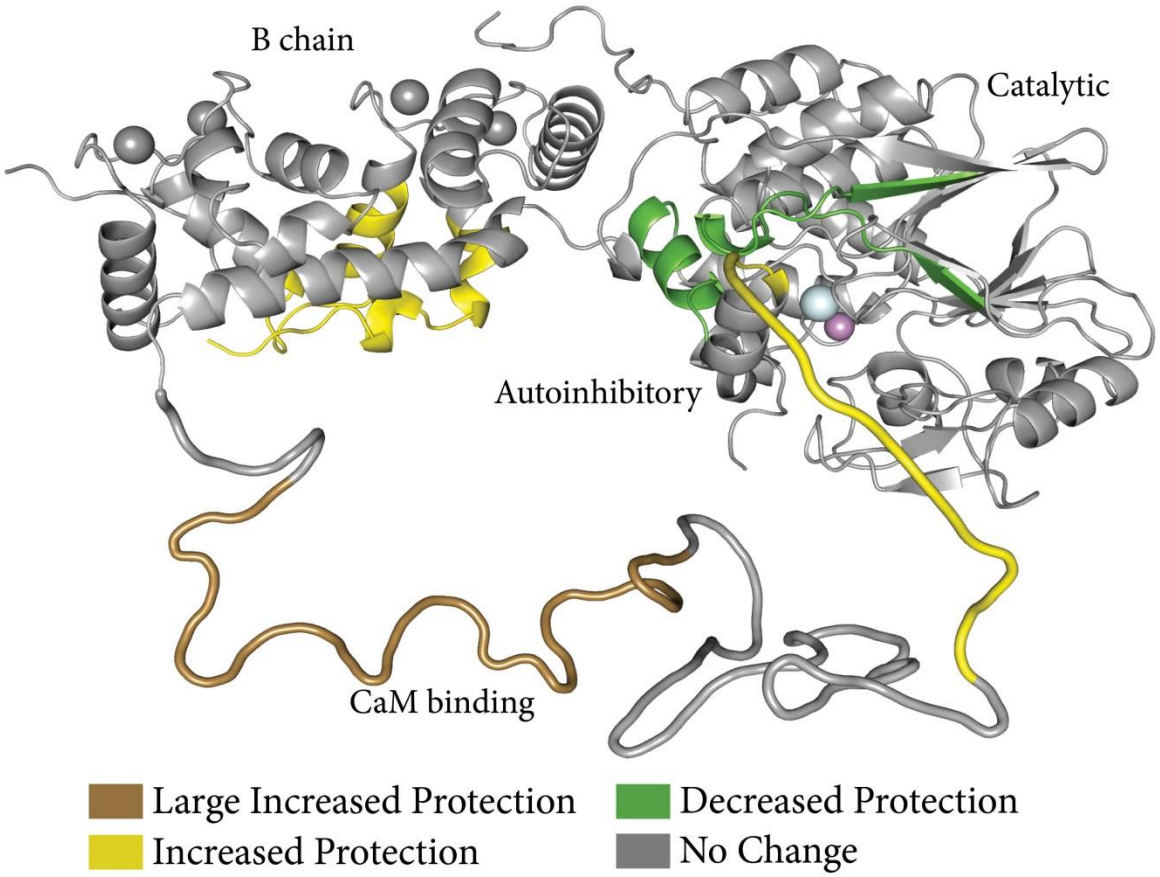


Figure 2.8. Effects of calmodulin on calcineurin as measured by hydrogen exchange mass spectrometry displayed on the modified crystal structure of CaN (PDB 1AUI).[16] The changes in protection are colored as shown in legend and are derived from the differential exchange data shown in Figure 2.7.

2.3.2.1 CaM binds to CaN at the regulatory domain

To evaluate the local conformational impact by CaM binding, we compared the rate of HX of the free CaN and bound CaN states of the calmodulin-binding domain region. The deuterium uptake curves for all the peptides are presented in Figure 2.S2. Representative deuterium uptake data by a peptide from calmodulin-binding domain is shown in Figure 2.6 (A 391-406). All the peptides located in the calmodulin-binding domain region exchange very rapidly in the absence of CaM and the exchanges were complete within fastest time point (13s),

indicating that the calmodulin-binding domain is unstructured in the absence of CaM. In contrast, when CaM binds the calmodulin-binding domain, the patterns of exchange by the peptides in this region become quite different. The rate of exchange becomes significantly slower after CaM binding and the peptides become strongly protected even after 24 h of exposure to D-$_2$O. We observed the largest increase in protection in the calmodulin-binding domain (see Figure 2.7). The strongest protection was –53 Da. The overlapping peptides covering the calmodulin-binding domain region are A 378-406, A 382-406, A 391-406, and A 407-413 (corresponding to peptide numbers 77-80 in Figure 2.7a, also see Figure 2.S2). In addition, CaM binding caused one other peptide in the regulatory domain (A 455-470 corresponding to peptide 91 in Figure 2.7a), which is distant from the calmodulin-binding domain region, to exchange slightly slower than the free form of CaN. These are the only two regions in the regulatory domain where changes in protection from exchange when CaM binds full length CaN were observed.

Unlike our previous work with CaM binding to the isolated RD-AID-CT,[(21)] in full length CaN, CaM binding did not cause statistically significant changes in protection from exchange in the peptides in the region adjacent to the C-terminal side of the calmodulin-binding domain, residues 414 to 450 (corresponding to peptide numbers 81-90 in Figure 2.7a).

2.3.2.2 Allosteric effects of CaM binding

In addition to the calmodulin-binding domain conformational change, calmodulin binding led to allosteric changes in CaN in locations distant from the binding site. Calmodulin binding caused several regions in the catalytic domain of CaN to exchange differently. These peptides are located near CaN's active site. The HX kinetics of one of the segments located in this region (A 303-311) is shown in Figure 2.6, while data for the other peptides are shown in Figure 2.S2. The peptides covering the area near the CaN's active site are A 160-176, A 303-311, A 303-312,

and A 303-314 (corresponding to peptide numbers 34, 63-65 in Figure 2.7a). The deuteration percentages of these peptides in both states of CaN were very low (15-30%) even after 24 h of exposure to D_2O. Although these peptides are located on the surface of the catalytic domain, proximity to the surface of the protein does not necessarily produce high deuteration.[38] The peptides located in the area near CaN's active site region exchange slightly more rapidly in the presence of CaM, indicating that this region become less protected after calmodulin binding. The strongest decrease in protection in the catalytic domain was 2.1 Da. There were no statistically significant differences in the other peptides located in the catalytic domain including CaN B-binding domain.

2.3.2.3 Effects of CaM binding on CaN B chain

Another region that undergoes altered hydrogen exchange after CaM binding is in the CaN B chain (CaN B). The HX kinetics of one of the segments located in this region (B 153-169) is shown in Figure 2.6, while data for the other peptides are shown in Figure 2.S2. We observed a significant increase in protection in peptide segments located in the C-terminal region of CaN B in the presence of calmodulin. The peptides covering the C-terminal region of CaN B are B 153-169, B 155-169, B 160-169, and B 161-169 (corresponding to peptide numbers 36-39 in Figure 2.7b). The strongest protection for this region was −5.4 Da. For the other regions of CaN B, there were no statistically significant changes in the deuterium uptake data.

2.3.2.4 CaM binding does not induce any changes in hydrogen exchange by the autoinhibitory domain

The peptides covering the AID region showed no changes in protection from exchange. These peptides are A 457-470, A 471-496, and A 472-496 (corresponding to peptide number 92-94 in Figure 2.7a, also see Figure 2.S2). Representative deuterium uptake data by the peptide A

471-496 are shown in Figure 2.6. The hydrogen exchange by the peptides located in this region reached a plateau within fastest time point (13 s) in the absence of CaM. When calmodulin bound the calmodulin-binding domain, no changes in the deuterium uptake by the peptides were observed.

2.4 Discussion

To explain the mechanism of CaN activation by CaM binding, we need to have a good understanding of the detailed structure and dynamics of CaN in solution in both states: free and bound to CaM. The previous models of CaN activation were either based on studies using truncated region of CaN or based on crystallographic studies using full-length CaN. In such cases, the models of activation may not reliably represent how full-length CaN undergoes conformational change in solution. In addition, some of the models were only based on low resolution biochemical and biophysical techniques. Our main goal in this study was to understand the mechanism of CaN activation in solution at higher resolution. For this purpose, we have used HX mass spectrometry to determine the CaN conformation at high calcium concentrations in the absence and presence of its binding partner, CaM, using full-length CaN. Here, we discuss the most widely accepted model of CaN activation, what we observed with the HX mass spectrometry data, and finally propose a new model of activation.

There is limited information about CaN conformation at no or low calcium concentrations and absence of CaM. The available information about CaN in this state is based on limited proteolysis and some other biochemical studies. Based on limited proteolysis of CaN in the absence of calcium, the calmodulin-binding domain and AID of the CaN regulatory domain are protected against proteolytic attack, indicating these two regions of the regulatory domain may remain packed against the CaN catalytic domain.[(5)] At high calcium concentration,

calcium binds to the low-affinity sites on CaN B, resulting in dissociation of the calmodulin-binding domain from the CaN catalytic domain and the protein becomes partially active.[15] The crystal structure of full-length CaN at high calcium concentration in the absence of calmodulin was first resolved by Kissinger et al. (see Figure 2.1, PDB: 1AUI).[16] In this structure, the electron density for the major part of regulatory domain including calmodulin-binding domain and C-terminal side of the regulatory domain is missing, indicating the disordered nature of these domains in the presence of excess of calcium in the absence of CaM. Fourier transform infrared spectroscopy and HX mass spectrometry data for the isolated RD-AID-CT construct also showed that the regulatory domain is disordered in the absence of CaM.[21, 39] A crystal structure of full-length CaN at high calcium concentration in the presence of calmodulin is not available to describe the conformational transition from a partially active form to the fully active form. Based on biochemical and biophysical techniques using both full-length CaN and truncated CaN constructs,[21, 40, 41] a model for CaN activation in the presence of calcium/calmodulin has been proposed and described as follows. At high calcium concentration, CaM binds four calcium ions and then subsequently binds to the calmodulin-binding domain of CaN. This binding causes a significant ordering in the calmodulin-binding domain and a weak ordering in a 38-residue region C-terminal to the calmodulin-binding domain of the regulatory domain. It has previously proposed that this ordering may cause displacement of the AID from the CaN active site leading to fully activated CaN.[21] Several techniques including HX, circular dichroism, and fluorescence anisotropy for the study have been used. The previous work supported the accepted mechanism of CaN activation. Since the previous HX data was mostly based on data obtained from the isolated RD-AID-CT, we wanted to extend the HX work using full-length CaN. The goal in the present work was to understand the conformational transition from partially active to fully active.

For calmodulin-binding domain, we observe that calmodulin-binding domain becomes fully deuterated within fastest labeling time point, indicating that calmodulin-binding domain might be unstructured in the absence of CaM. It seems that this region is very flexible and lacks detectable stable secondary structure in the absence of CaM (see Figure 2.6 and 2.7a). The crystal structure of CaN is missing this region, indicating that this region is unstructured in the absence of CaM.[16] But when CaM binds CaN regulatory domain, the calmodulin-binding domain becomes strongly protected from HX, indicating that this region becomes structured (see Figure 2.6, 2.7a, 2.8, and 2.S1). This structuring is expected to be an α-helix in the calmodulin-binding domain.[21, 42] This kind of structuring happens with many other proteins contain calmodulin-binding domain.[20, 43, 44] Similarly, using Fourier transform infrared spectroscopy to determine the secondary structure of CaN in the presence and absence of CaM, Zhao et al. demonstrated that the regulatory domain of CaN undergoes a large-scale folding.[26] Thus, these HX results are consistent with the accepted models regarding calmodulin-binding domain.

The HX data for the adjacent region to the calmodulin-binding domain, however, look different than what was previously observed for this region. Previous HX data on the isolated regulatory domain fragment of CaN demonstrated that there is weak protection in the region adjacent to the C-terminal region of calmodulin-binding domain.[21] In full-length CaN, however, we did not observe any significant protection from exchange in the weakly-protected region on the C-terminal side of the calmodulin-binding domain after CaM binding (see Figure 2.6, 2.7a, and 2.8). The rate of exchange by the peptides in this region was very rapid in both the presence and absence of CaM (see Figure 2.6 and 2.S2). Our data from full-length CaN do not agree with the previously published HX data obtained from the isolated regulatory domain fragment of CaN in this region. Based on the full-length CaN HX data in the presence and absence of CaM, we

conclude that the flexibility of the region immediately next to the C-terminal site of the calmodulin-binding domain remains unaffected by CaM binding. Our data suggest that this region does not engage with CaM as previously proposed. The reason could be because CaM may either prefer interacting with another region of CaN B which did not exist in the isolated regulatory domain experiment or because CaM may prefer interacting with the region N-terminal to the AID, or both.

Full-length CaN HX data also demonstrated that another region in the regulatory domain immediately N-terminal to the AID becomes weakly protected after CaM binding (A 455-470) (see Figure 2.6 and peptide 91 in Figure 2.7a), but such protection was not observed in the isolated regulatory domain fragment of CaN.[21] Other work also suggests that the region adjacent to the N-terminal region of AID has a role in regulating the CaN activation. Perrino et al. have shown that there is 38-residue region adjacent to the N-terminal of AID that has a role in the activation of CaN.[41] Using enzyme activity assays, they found that the deletion of the AID, but not the regulatory domain, caused CaN to become partially active. Full activation of CaN was only observed when the 38-residue region and the AID were deleted.[22, 41] Based on the activity assays data and our HX data, the region adjacent to the N-terminal region of AID might have a role in CaN activation and there might be an interaction between this region and CaM, but the interaction is likely not very stable as indicated by only a weak protection from exchange (see Figure 2.7a and 2.8). Based on different HX results obtained for isolated regulatory domain fragment versus full-length CaN, our data lead us to conclude that HX studies on truncated region of CaN may not accurately represent some aspects of the conformational changes that CaM binding induces in the whole protein.

The HX data from full-length CaN demonstrate that the CaM binding leads to allosteric effects in the area near active site in the catalytic domain. In contrast to the other regions of CaN, this region is the only region of CaN to become more flexible after CaM binding to CaN. A slight decrease in protection from exchange was observed in the peptides located in the regions near the active site (see Figure 2.6, 2.7a, and 2.8). The allosteric effects in this region appear to have a role in regulating the CaN activations because this region may be directly affected when CaN become fully active and ease the binding of CaN to its substrate. Previous studies showed that there is a relation between the active site of enzymes and their flexibility.[45-47] It is been suggested that the enzyme active sites need to be flexible to ease the conformational change during activation or inactivation. Although there is a change in the HX by the peptides located near the active site of the catalytic domain, the deuteration percentages of these peptides were very low (15-30%). Since these peptides are located on the surface of the catalytic domain, they might be expected to exchange rapidly. However in contrast to this assumption, the proximity to the surface of the protein does not necessarily cause rapid exchange.[38, 48] These peptides might have slowly exchanging amide hydrogens that are hydrogen bonded to hydrogens of the same helix or hydrogen bonded to a defined water molecules that also hydrogen bond to main chain carbonyls.

Another region that undergoes altered hydrogen exchange is CaN B. We found that the peptides located in the C-terminal region of CaN B become significantly protected after CaM binding. Other regions of CaN B remain unaffected by CaM binding (see Figure 2.6, 2.7b, 2.8, and 2.S1). Previous kinetic analysis demonstrated that the presence of CaN B was required for CaN activation.[22] Based on our data, one possible interpretation is that the protection in the C-terminal region of CaN B is a result of an allosteric effect of CaM binding. One other

interpretation is that there is an interaction between CaN B and calmodulin. We favor the interpretation of the binding between CaN B and calmodulin because CaN B may regulate the CaN activation by anchoring the calmodulin to the calmodulin-binding domain. From the folded conformer of calmodulin bound to the calmodulin-binding domain peptide (PDB: 4Q5U),[(24)] we learn that calmodulin is about the same size as CaN B chain. Examination of the crystal structure of CaN free form (see Figure 2.1b), suggests that there is about the right amount of space for CaM to fit into the space between the catalytic domain and CaN B to keep the CaN in the active form. One example of similar conformation is the conformation of CaN bound to the drug-protein complex cyclosporine A-cyclophilin (PDB: 1MF8).[(14)] In the crystal structure, the drug-protein complex also fills the space between the catalytic domain and CaN B. Although the CaN/drug/protein complex is an inhibitory interaction in which some details are different than CaN:CaM binding, they might have a similar conformation. Therefore, we suggest that the interaction between the C-terminal region of CaN B and CaM and on the other hand, the interaction between the region immediately N-terminal to the AID and CaM may favor to form the space-occupied conformation of the CaN:CaM complex and cooperatively activate CaN.

The most interesting aspect of the full-length CaN HX data is what we found for the AID. We do not see changes in deuterium uptake by the peptides located in the AID region after CaM binding (see Figure 2.6, 2.7a, 2.8, and 2.S1), indicating that there is no structural change in the AID. The peptides in the AID were already about 50% deuterated at the fastest measured exchange time and remained at the same percentage even after 24 hours labeling. Based on our data, there are two possible interpretations to explain rapid, but low deuteration for the peptides in the AID region. One possible interpretation is that some residues of the AID are highly protected. The other possible interpretation is that the AID region undergoes extensive back-

exchange. To investigate which interpretation might be correct, we compared back exchange percentages of the AID peptides from full length CaN experiment with the same peptides from the RD-AID-CT construct isolated from full length CaN experiment. Since there is no active site in the isolated form of RD-AID-CT, there is nothing for the AID to bind to. Therefore, the back exchange percentage comparison would tell us whether the low deuteration of the AID peptides in full length of CaN is either due to extensive back-exchange or structural protection. We observed that the back exchange for the AID peptides in the isolated form of RD-AID-CT is also about 50-55%. The back exchange percentage is similar to the back exchange percentage observed for the AID peptides in the full-length CaN experiment. Thus, we prefer the second interpretation: that the AID peptides undergo extensive back-exchange under quench conditions. Based on our HX data for the full-length CaN, the AID is either unstructured and has no interaction with the catalytic domain or is unstructured and has interaction with the catalytic domain only through side chains. The HX experiment is not able to distinguish between these two possible states of the AID. Thus, based on our data, we can state that the AID is unstructured, not folded into the active site of CaN, and CaM binding does not change the HX of AID. This observation does not agree with the widely accepted model of CaN activation. Several contradictory results were reported about the AID and its role in the CaN activation. For example, Wang et al. reported that the activity of truncated form of CaN, which contains the calmodulin-binding domain (residue 1- 420), is higher than the native enzyme in the presence of calcium/CaM.[(40)] In another activity assay study, however, a similar activities were reported for the same truncation of CaN and the native full length CaN.[(39)] In a separate study, Ye et al. reported a crystal structure of CaN covalently linked to CaM which exhibits full CaM-activated activity (PDB: 4IL1).[(25)] Although CaM was covalently linked to CaN, CaM does not appear in

the crystal structure. In the study, no conformational change of the AID was observed, indicating that the AID was not released from the active site. It has also been suggested that the AID does not only include the region (468-490) as previously thought, but might also include other regions of the regulatory domain. Based on these results, the idea of reclassification of the AID has been suggested.[25] Based on our HX data from full-length CaN, the AID is unstructured and could be bound to the catalytic domain segments through side-chain contacts. If the AID associated with the catalytic domain only through side chain interactions, we would not know that these had been lost when CaM binds. The interaction would be invisible by HX for the AID. However, decreased protection in the catalytic domain segments could be explained by release or reorientation of an unstructured AID that only makes side-chain contacts.

The crystal structure of the β CaN isoform was recently solved to characterize the mechanism of CaN activation.[49] Based on biochemical studies, CaM does not remove the AID from the active site. The study also reports a structure of full length β CaN isoform which reveal an autoinhibitory segment (AIS) within the regulatory domain of CaN A. In the absence of CaM, this segment appears to be packed into a hydrophobic intersubunit (CaN A and B) groove. In contrast, in the crystal structure of α CaN isoform (see Figure 2.1b), there is no evidence of the so-called AIS present in the structure. It has been suggested that the AIS is untraceable in α CaN due to AIS's high flexibility.[49] Based on our HX data, we conclude that the AIS (A 407-414) is very flexible and unstructured in the absence of CaM as this region becomes fully deuterated within fastest labeling time point. It is possible that the AIS exists in two conformations. If this were the case, we might see EX1 kinetics, which produces bimodal isotopic distributions in the mass spectra,[50, 51] in the AIS. However, we do not see evidence of EX1 kinetics in the AIS that

would support the existence of two conformations of AIS in α and β CaN might represent snapshots of the ensemble conformations (see Figure 2.S3).

In summary, based on HX results obtained from full-length αCaN presented here in this study, we propose a model for the mechanism of CaN activation as shown in Figure 2.9. Since we ran our HX experiment in the presence of high calcium concentration, our model of activation is only focused on the structural changes in the presence or absence of CaM. The low- and high-affinity sites on CaN B and CaM are already occupied with calcium to stimulate the basal activity of CaN in the absence of CaM. In the absence of CaM, the AID is unstructured and bound to the active site through side chains and the other part of regulatory domain including calmodulin-binding domain and CT are largely unstructured. In the presence of CaM, a series of conformational changes, both near and distant from the active site, are responsible for regulation of CaN activity. The CaM binding causes calmodulin-binding domain to form an α-helix in this region. The C-terminal region of CaN B and the region immediately N-terminal to the AID interact with CaM to anchor CaM to the binding site. The CaN:CaM binding causes the area near the active site to become slightly more flexible. One possible explanation for the slight increase in flexibility in this region might be because of reorientation of the AID. Our model of CaN activation is based on data obtained from a technique that can provide molecular details with high spatial resolution at the peptide-level to probe conformational changes in solution. Our model challenges the currently available models of CaN activation. To investigate the complex multistep regulatory mechanism of CaN activation and have a complete story of the activation, future studies on CaN in the absence of calcium in solution will be required.

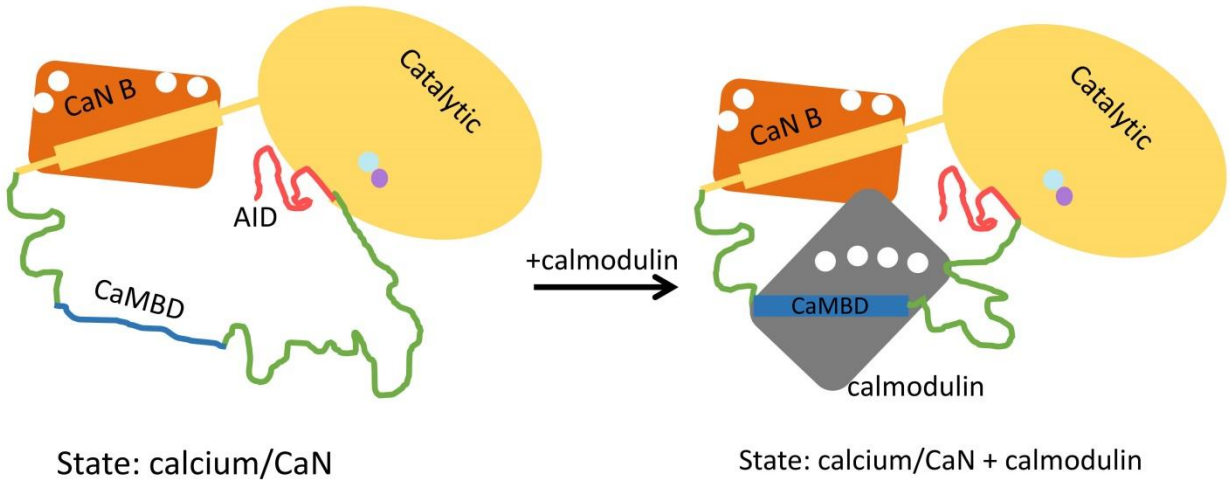


Figure 2.9. Proposed model of calcineurin activation after calmodulin binding based on full-length calcineurin data. The proposed model based on data obtained by HX-MS for two states of CaN: CaN at high calcium concentration in the absence of CaM shown in the left panel of the figure and CaN at high calcium concentration in the presence of CaM shown in the right panel of the figure. The open dots represent calcium ions bound to CaN B and calmodulin. The purple and light blue dots in the catalytic domain represent zinc and iron ions, respectively.

2.5 References

1. Wang, J. H., and Desai, R. (1976) A brain protein and its effect on the Ca2+-and protein modulator-activated cyclic nucleotide phosphodiesterase, *Biochemical and biophysical research communications 72*, 926-932.

2. Aramburu, J., Rao, A., and Klee, C. B. (2000) Calcineurin: from structure to function, *Current topics in cellular regulation 36*, 237-295.

3. Rusnak, F., and Mertz, P. (2000) Calcineurin: form and function, *Physiological reviews 80*, 1483-1521.

4. Li, H., Rao, A., and Hogan, P. G. (2011) Interaction of calcineurin with substrates and targeting proteins, *Trends in cell biology 21*, 91-103.

5. Yang, S.-A., and Klee, C. B. (2000) Low Affinity Ca2+-Binding Sites of Calcineurin B Mediate Conformational Changes in Calcineurin A, *Biochemistry 39*, 16147-16154.

6. Macian, F. (2005) NFAT proteins: key regulators of T-cell development and function, *Nature reviews. Immunology 5*, 472-484.

7. Mancini, M., and Toker, A. (2009) NFAT proteins: emerging roles in cancer progression, *Nature reviews. Cancer 9*, 810-820.

8. Crabtree, G. R., and Olson, E. N. (2002) NFAT signaling: choreographing the social lives of cells, *Cell 109 Suppl*, S67-79.

9. Ermak, G., Morgan, T. E., and Davies, K. J. (2001) Chronic overexpression of the calcineurin inhibitory gene DSCR1 (Adapt78) is associated with Alzheimer's disease, *The Journal of biological chemistry 276*, 38787-38794.

10. Hoeffer, C. A., Dey, A., Sachan, N., Wong, H., Patterson, R. J., Shelton, J. M., Richardson, J. A., Klann, E., and Rothermel, B. A. (2007) The Down syndrome critical region protein RCAN1 regulates long-term potentiation and memory via inhibition of phosphatase signaling, *The Journal of neuroscience 27*, 13161-13172.

11. Wang, P., and Heitman, J. (2005) The cyclophilins, *Genome biology 6*, 226.

12. Griffith, J. P., Kim, J. L., Kim, E. E., Sintchak, M. D., Thomson, J. A., Fitzgibbon, M. J., Fleming, M. A., Caron, P. R., Hsiao, K., and Navia, M. A. (1995) X-ray structure of calcineurin inhibited by the immunophilin-immunosuppressant FKBP12-FK506 complex, *Cell 82*, 507-522.

13. Huai, Q., Kim, H. Y., Liu, Y., Zhao, Y., Mondragon, A., Liu, J. O., and Ke, H. (2002) Crystal structure of calcineurin-cyclophilin-cyclosporin shows common but distinct recognition of immunophilin-drug complexes, *Proceedings of the national academy of sciences of the United States of America 99*, 12037-12042.

14. Jin, L., and Harrison, S. C. (2002) Crystal structure of human calcineurin complexed with cyclosporin A and human cyclophilin, *Proceedings of the national academy of sciences of the United States of America 99*, 13522-13526.

15. Klee, C. B., Ren, H., and Wang, X. (1998) Regulation of the calmodulin-stimulated protein phosphatase, calcineurin, *The Journal of biological chemistry 273*, 13367-13370.

16. Kissinger, C. R., Parge, H. E., Knighton, D. R., Lewis, C. T., Pelletier, L. A., Tempczyk, A., Kalish, V. J., Tucker, K. D., Showalter, R. E., Moomaw, E. W., and et al. (1995) Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex, *Nature 378*, 641-644.

17. Manalan, A. S., and Klee, C. B. (1983) Activation of calcineurin by limited proteolysis, *Proceedings of the national academy of sciences of the United States of America 80*, 4291-4295.

18. Hashimoto, Y., Perrino, B. A., and Soderling, T. R. (1990) Identification of an autoinhibitory domain in calcineurin, *The Journal of biological chemistry 265*, 1924-1927.

19. Watanabe, Y., Perrino, B. A., Chang, B. H., and Soderling, T. R. (1995) Identification in the calcineurin A subunit of the domain that binds the regulatory B subunit, *The Journal of biological chemistry 270*, 456-460.

20. Crivici, A., and Ikura, M. (1995) Molecular and structural basis of target recognition by calmodulin, *Annual review of biophysics and biomolecular structure 24*, 85-116.

21. Rumi-Masante, J., Rusinga, F. I., Lester, T. E., Dunlap, T. B., Williams, T. D., Dunker, A. K., Weis, D. D., and Creamer, T. P. (2012) Structural basis for activation of calcineurin by calmodulin, *Journal of molecular biology 415*, 307-317.

22. Perrino, B. A., Ng, L. Y., and Soderling, T. R. (1995) Calcium regulation of calcineurin phosphatase activity by its B subunit and calmodulin. Role of the autoinhibitory domain, *The Journal of biological chemistry 270*, 7012.

23. Hubbard, M. J., and Klee, C. B. (1989) Functional domain structure of calcineurin A: mapping by limited proteolysis, *Biochemistry 28*, 1868-1874.

24. Dunlap, T. B., Guo, H. F., Cook, E. C., Holbrook, E., Rumi-Masante, J., Lester, T. E., Colbert, C. L., Vander Kooi, C. W., and Creamer, T. P. (2014) Stoichiometry of the calcineurin regulatory domain-calmodulin complex, *Biochemistry 53*, 5779-5790.

25. Ye, Q., Feng, Y., Yin, Y., Faucher, F., Currie, M. A., Rahman, M. N., Jin, J., Li, S., Wei, Q., and Jia, Z. (2013) Structural basis of calcineurin activation by calmodulin, *Cellular signalling 25*, 2661-2667.

26. Zhao, Y., Yang, H., Meng, K., and Yu, S. (2014) Probing the Ca2+/CaM-induced secondary structural and conformational changes in calcineurin, *International journal of biological macromolecules 64*, 453-457.

27. Ye, Q., Li, X., Wong, A., Wei, Q., and Jia, Z. (2006) Structure of calmodulin bound to a calcineurin peptide: a new way of making an old binding mode, *Biochemistry 45*, 738-745.

28. Englander, S. W., and Kallenbach, N. R. (1983) Hydrogen exchange and structural dynamics of proteins and nucleic acids, *Quarterly reviews of biophysics 16*, 521-655.

29. Marcsisin, S. R., and Engen, J. R. (2010) Hydrogen exchange mass spectrometry: what is it and what can it tell us?, *Analytical and bioanalytical chemistry 397*, 967-972.

30. Charbonneau, H., Hice, R., Hart, R. C., and Cormier, M. J. (1983) Purification of calmodulin by Ca2+-dependent affinity chromatography, *Methods in enzymology 102*, 17-39.

31. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) Measurement of protein using bicinchoninic acid, *Analytical biochemistry 150*, 76-85.

32. Wang, L., Pan, H., and Smith, D. L. (2002) Hydrogen exchange-mass spectrometry: optimization of digestion conditions, *Molecular & cellular proteomics : MCP 1*, 132-138.

33. Busby, S. A., Chalmers, M. J., and Griffin, P. R. (2007) Improving digestion efficiency under H/D exchange conditions with activated pepsinogen coupled columns, *International journal of mass spectrometry 259*, 130-139.

34. Kavan, D., and Man, P. (2011) MSTools—Web based application for visualization and presentation of HXMS data, *International journal of mass spectrometry 302*, 53-58.

35. Houde, D., Berkowitz, S. A., and Engen, J. R. (2011) The utility of hydrogen/deuterium exchange mass spectrometry in biopharmaceutical comparability studies, *Journal of pharmaceutical sciences 100*, 2071-2086.

36. Quintana, A. R., Wang, D., Forbes, J. E., and Waxham, M. N. (2005) Kinetics of calmodulin binding to calcineurin, *Biochemical and biophysical research communications 334*, 674-680.

37. O'Donnell, S. E., Yu, L., Fowler, C. A., and Shea, M. A. (2011) Recognition of beta-calcineurin by the domains of calmodulin: thermodynamic and structural evidence for distinct roles, *Proteins 79*, 765-786.

38. Skinner, J. J., Lim, W. K., Bedard, S., Black, B. E., and Englander, S. W. (2012) Protein hydrogen exchange: testing current models, *Protein science : a publication of the Protein Society 21*, 987-995.

39. Shen, X., Li, H., Ou, Y., Tao, W., Dong, A., Kong, J., Ji, C., and Yu, S. (2008) The secondary structure of calcineurin regulatory region and conformational change induced by calcium/calmodulin binding, *The Journal of biological chemistry 283*, 11407.

40. Wang, H., Du, Y., Xiang, B., Lin, W., Li, X., and Wei, Q. (2008) A renewed model of CNA regulation involving its C-terminal regulatory domain and CaM, *Biochemistry 47*, 4461-4468.

41. Perrino, B. A. (1999) Regulation of calcineurin phosphatase activity by its autoinhibitory domain, *Archives of biochemistry and biophysics 372*, 159-165.

42. Dunlap, T. B., Cook, E. C., Rumi-Masante, J., Arvin, H. G., Lester, T. E., and Creamer, T. P. (2013) The distal helix in the regulatory domain of calcineurin is important for domain stability and enzyme function, *Biochemistry 52*, 8643-8651.

43. Vetter, S. W., and Leclerc, E. (2003) Novel aspects of calmodulin target recognition and activation, *European journal of biochemistry 270*, 404-414.

44. Zhang, Y., Tan, H., Chen, G., and Jia, Z. (2010) Investigating the disorder-order transition of calmodulin binding domain upon binding calmodulin using molecular dynamics simulation, *Journal of molecular recognition 23*, 360-368.

45. Zou, C. L. (2001) [Conformational flexibility of enzyme active sites], *Sheng li ke xue jin zhan [Progress in physiology] 32*, 7-12.

46. Otyepka, M., Berka, K., and Anzenbacher, P. (2012) Is there a relationship between the substrate preferences and structural flexibility of cytochromes P450?, *Current drug metabolism 13*, 130-142.

47. Deschamps, N., Simoes-Pires, C. A., Carrupt, P. A., and Nurisso, A. (2015) How the flexibility of human histone deacetylases influences ligand binding: an overview, *Drug discovery today 20*, 736-742.

48. Skinner, J. J., Lim, W. K., Bedard, S., Black, B. E., and Englander, S. W. (2012) Protein dynamics viewed by hydrogen exchange, *Protein science : a publication of the Protein Society 21*, 996-1005.

49. Li, S. J., Wang, J., Ma, L., Lu, C., Wang, J., Wu, J. W., and Wang, Z. X. (2016) Cooperative autoinhibition and multi-level activation mechanisms of calcineurin, *Cell research 26*, 336-349.

50. Konermann, L., Tong, X., and Pan, Y. (2008) Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches, *Journal of mass spectrometry 43*, 1021-1036.

51. Krishna, M. M., Hoang, L., Lin, Y., and Englander, S. W. (2004) Hydrogen exchange methods to study protein folding, *Methods (San Diego, Calif.) 34*, 51-64.

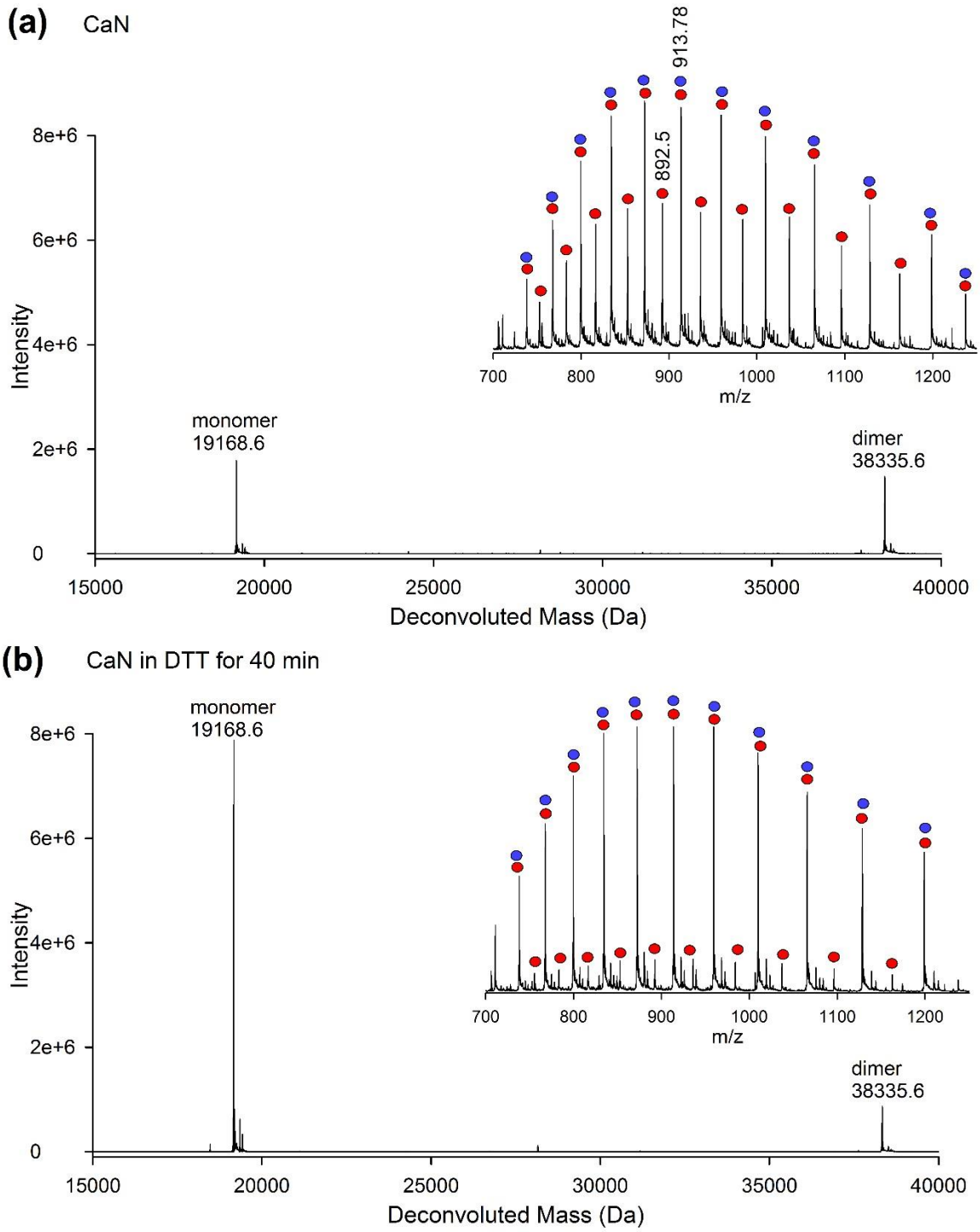
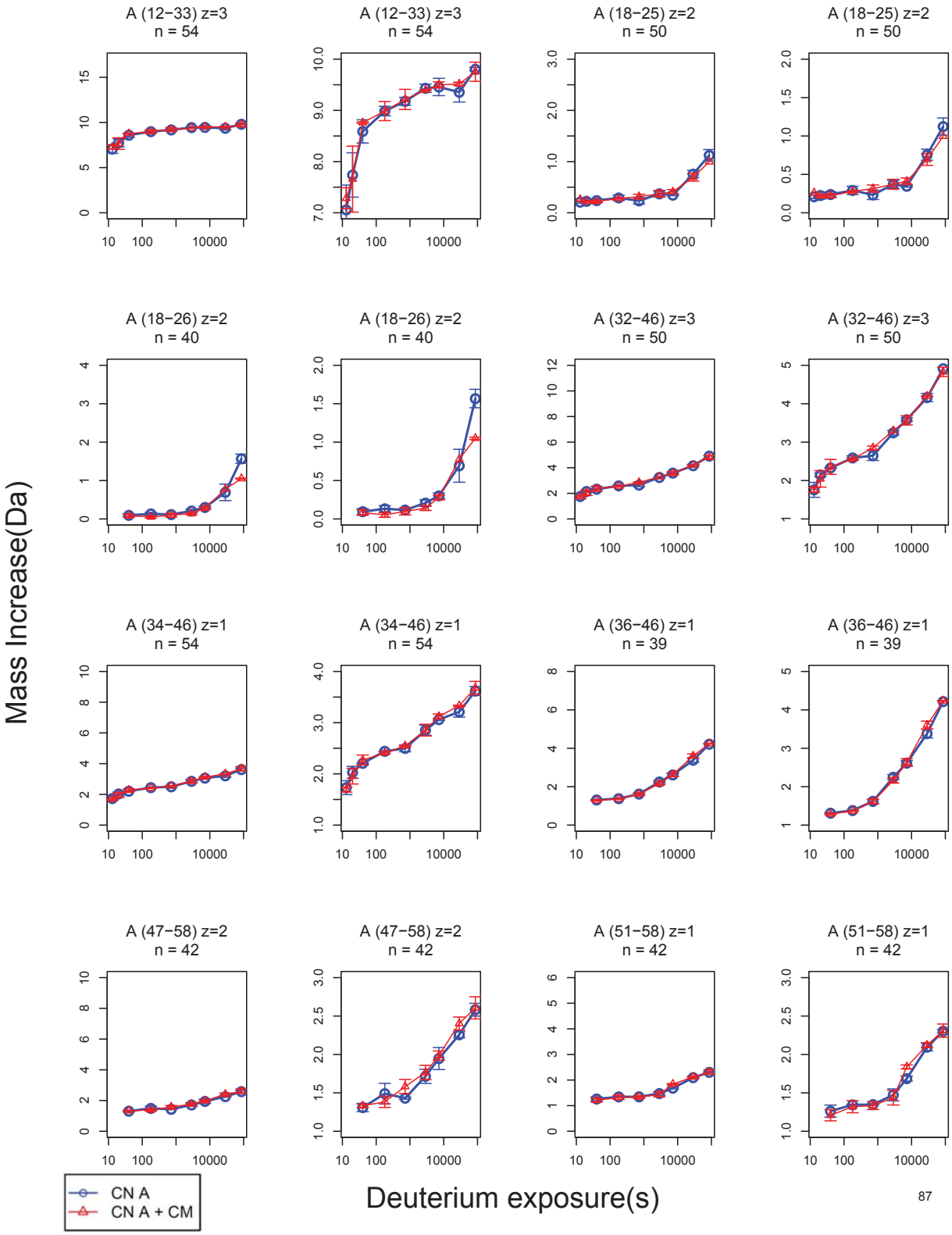
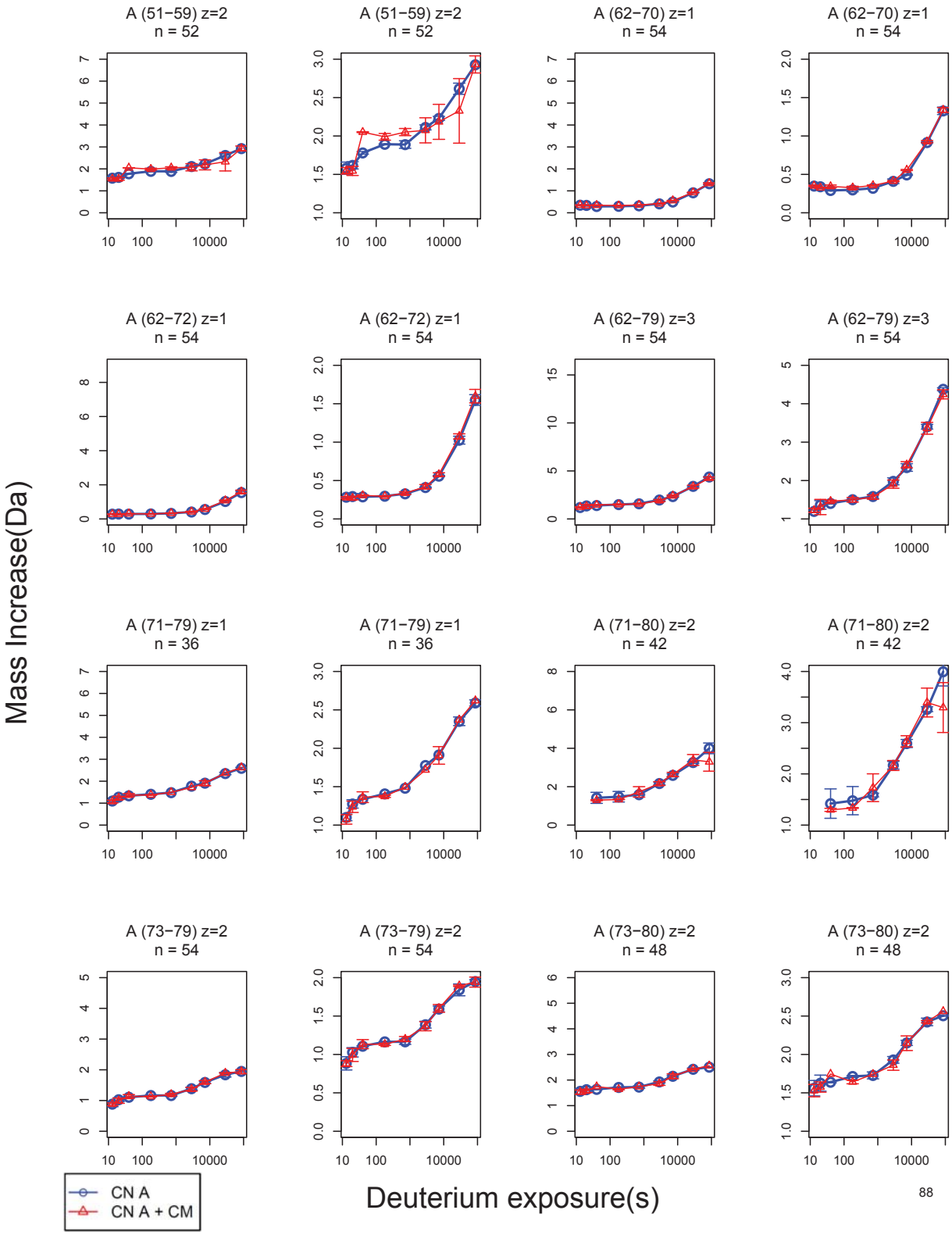
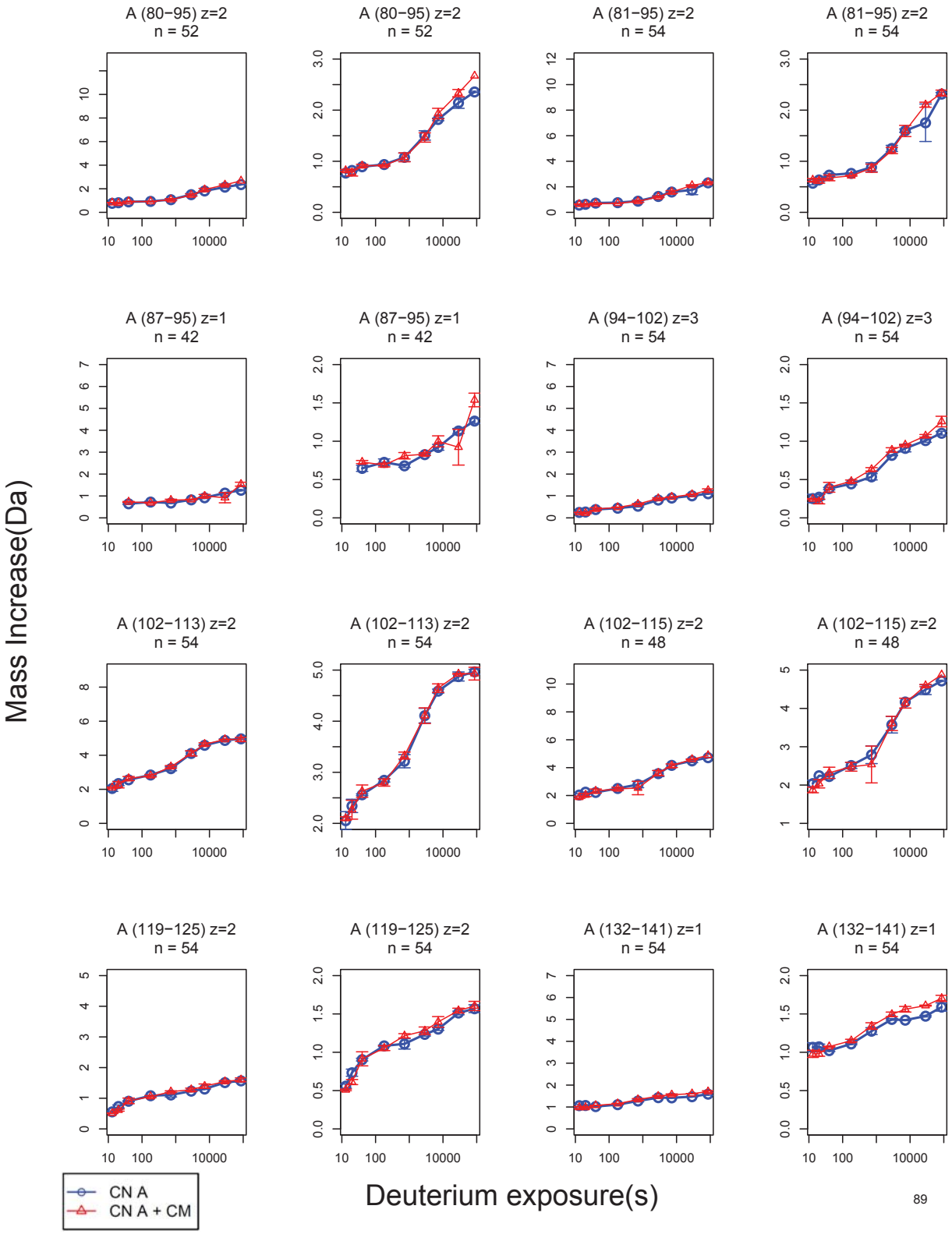
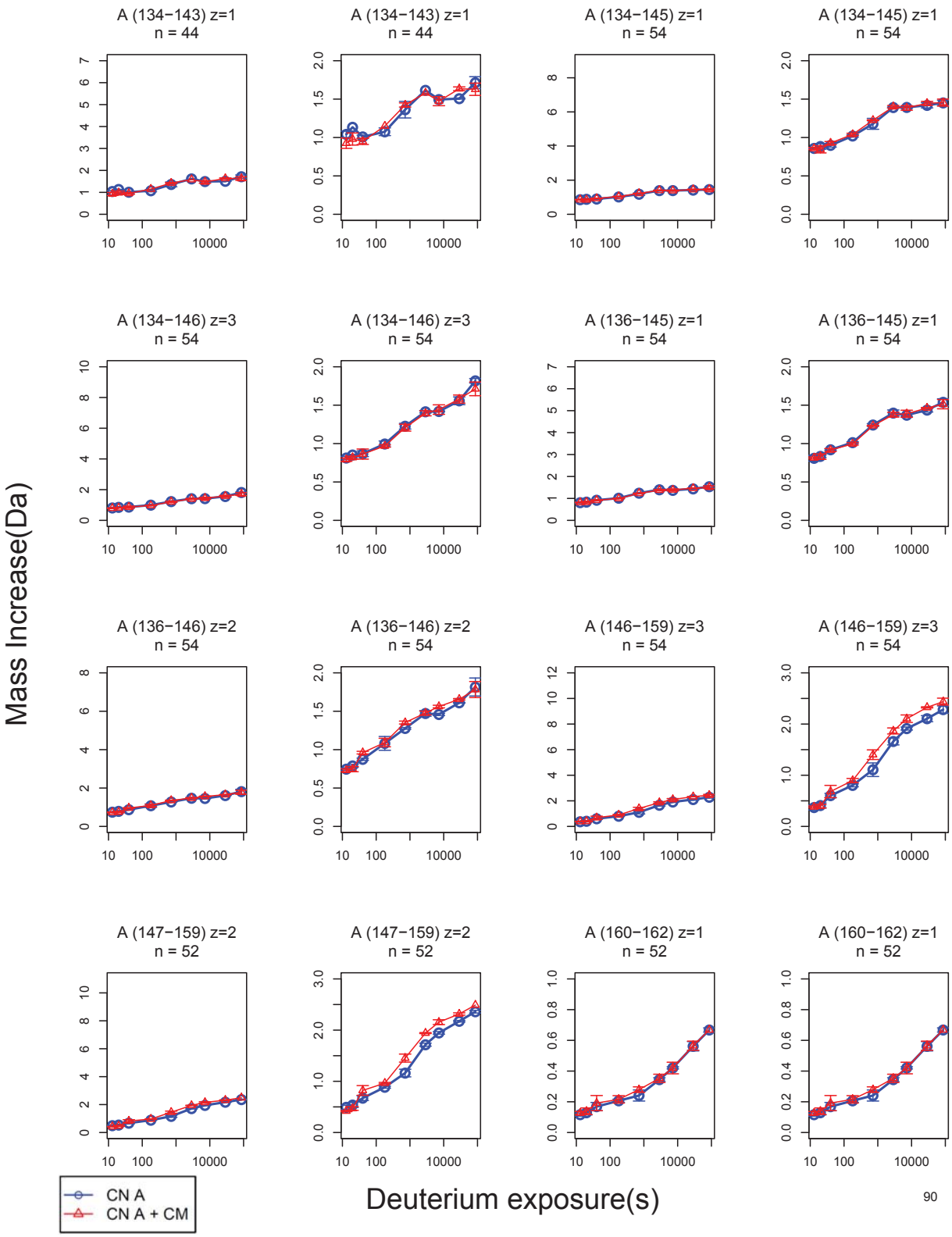


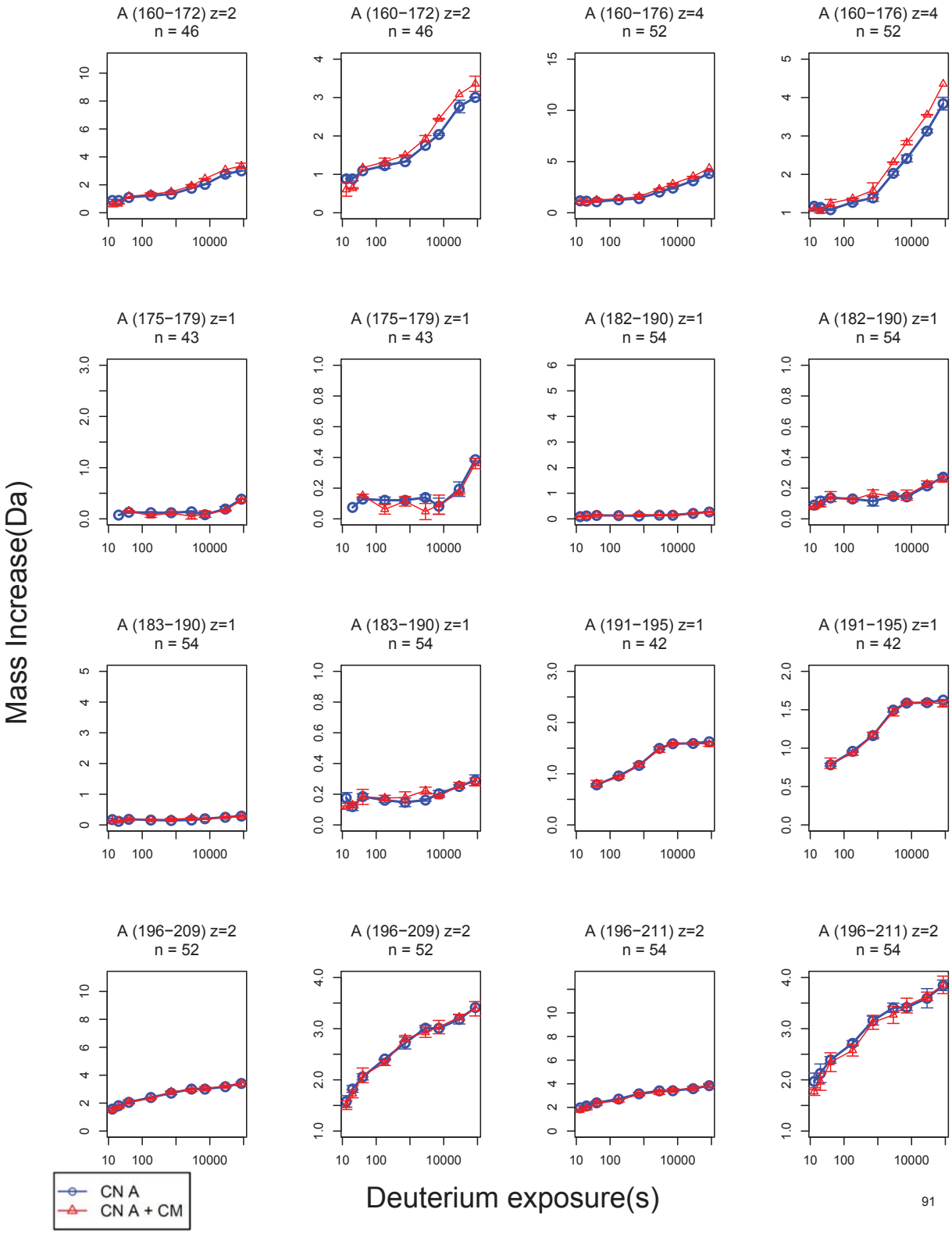
Figure 2.S1. Deconvoluted spectra and mass spectra (shown in the insets) of calcineurin B in (a) the absence of DTT and (b) in the presence of DTT after incubation for 40 min at 4 °C. In the inset figure (a), the m/z values of 913.78 and 892.50 correspond to monomer CaN B $[M+21H]^{21+}$ and dimer CaN B $[2M-2H+43H]^{43+}$, respectively. The CaN B monomer theoretical mass was 19168.7 Da and the dimer mass was 38335.4 Da (2×M – 2H). The intensity of the dimer decreases and the monomer intensity increase after incubation for 50 min in DTT.

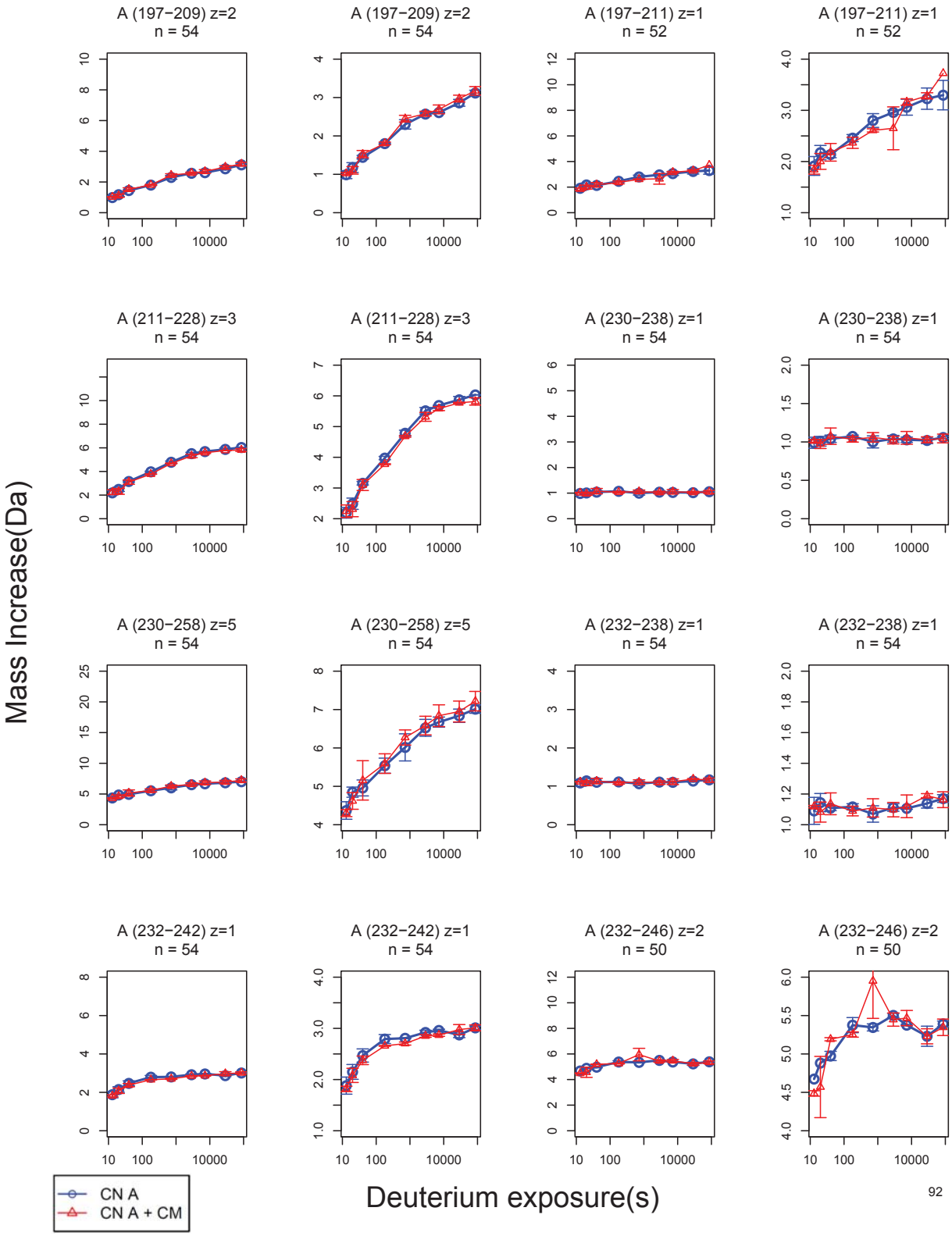

A (12−33) z=3
n = 54
A (12−33) z=3
n = 54
A (18−25) z=2
n = 50
A (18−25) z=2
n = 50
A (18−26) z=2
n = 40
A (18−26) z=2
n = 40
A (32−46) z=3
n = 50
A (32−46) z=3
n = 50
A (34−46) z=1
n = 54
A (34−46) z=1
n = 54
A (36−46) z=1
n = 39
A (36−46) z=1
n = 39
A (47−58) z=2
n = 42
A (47−58) z=2
n = 42
A (51−58) z=1
n = 42
A (51−58) z=1
n = 42
Mass Increase(Da)
Deuterium exposure(s)
CN A
CN A + CM

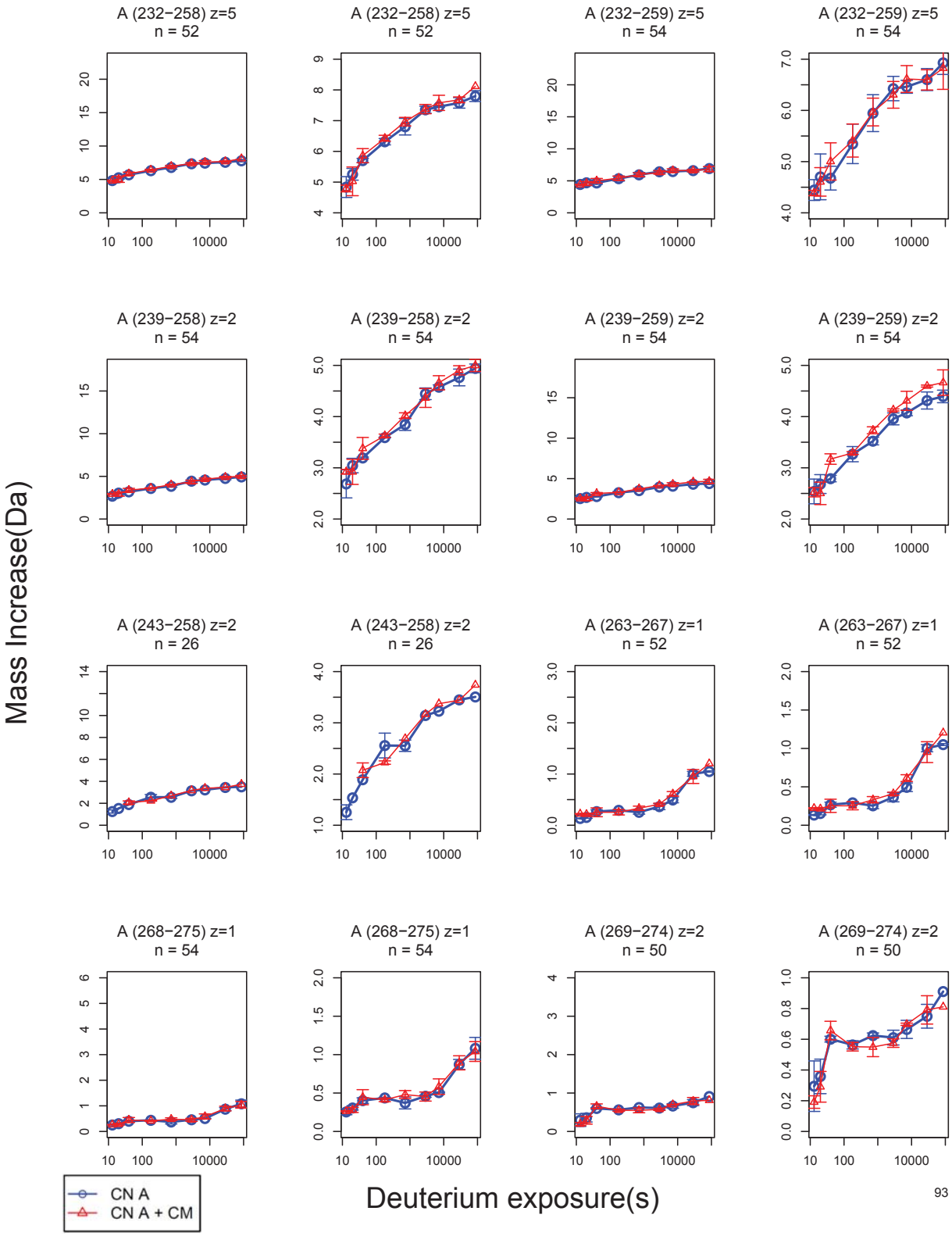

A (51−59) z=2
n = 52
A (51−59) z=2
n = 52
A (62−70) z=1
n = 54
A (62−70) z=1
n = 54
A (62−72) z=1
n = 54
A (62−72) z=1
n = 54
A (62−79) z=3
n = 54
A (62−79) z=3
n = 54
A (71−79) z=1
n = 36
A (71−79) z=1
n = 36
A (71−80) z=2
n = 42
A (71−80) z=2
n = 42
A (73−79) z=2
n = 54
A (73−79) z=2
n = 54
A (73−80) z=2
n = 48
A (73−80) z=2
n = 48
Mass Increase(Da)
Deuterium exposure(s)
CN A
CN A + CM

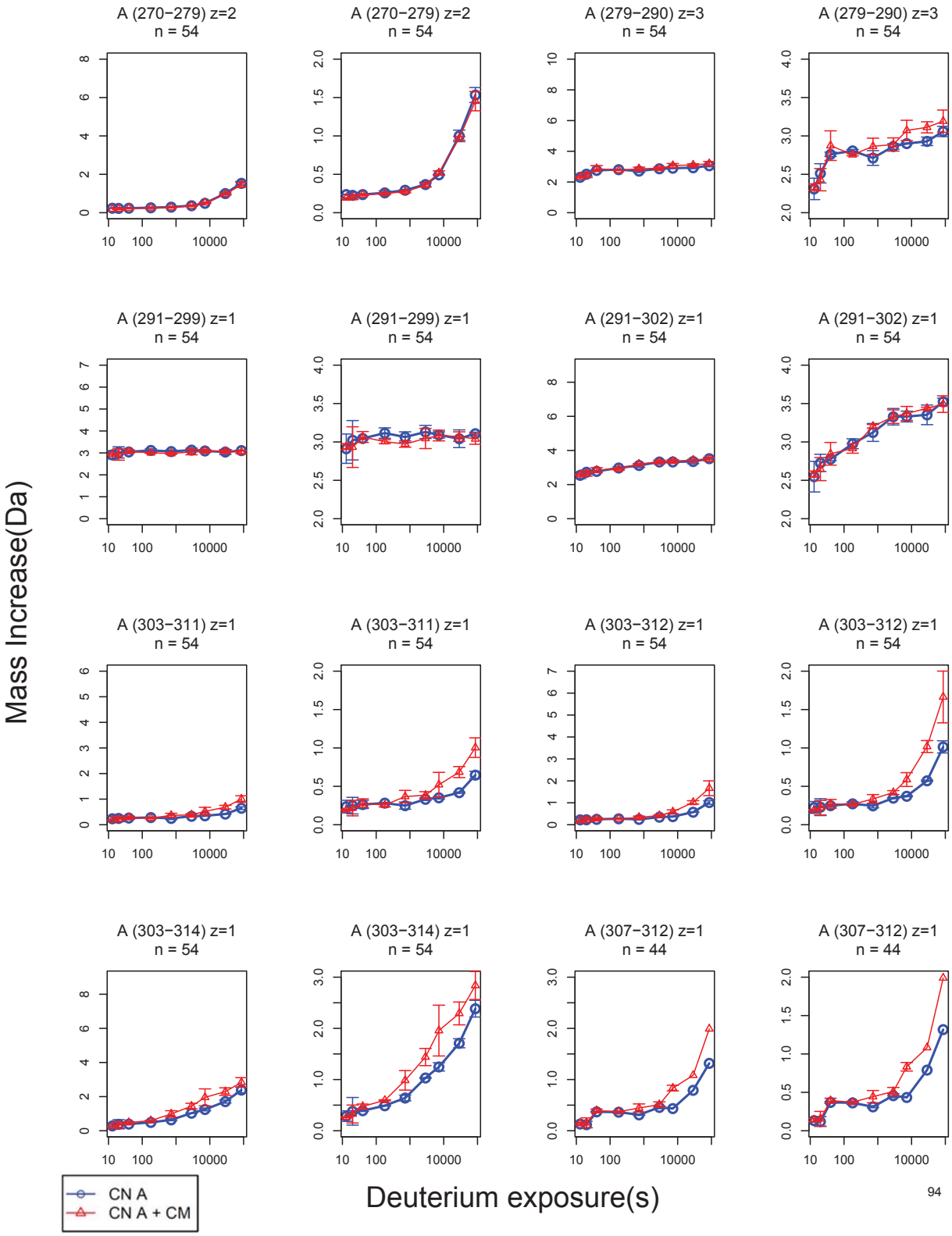

A (80−95) z=2
n = 52
A (80−95) z=2
n = 52
A (81−95) z=2
n = 54
A (81−95) z=2
n = 54
A (87−95) z=1
n = 42
A (87−95) z=1
n = 42
A (94−102) z=3
n = 54
A (94−102) z=3
n = 54
A (102−113) z=2
n = 54
A (102−113) z=2
n = 54
A (102−115) z=2
n = 48
A (102−115) z=2
n = 48
A (119−125) z=2
n = 54
A (119−125) z=2
n = 54
A (132−141) z=1
n = 54
A (132−141) z=1
n = 54
Mass Increase(Da)
Deuterium exposure(s)
CN A
CN A + CM

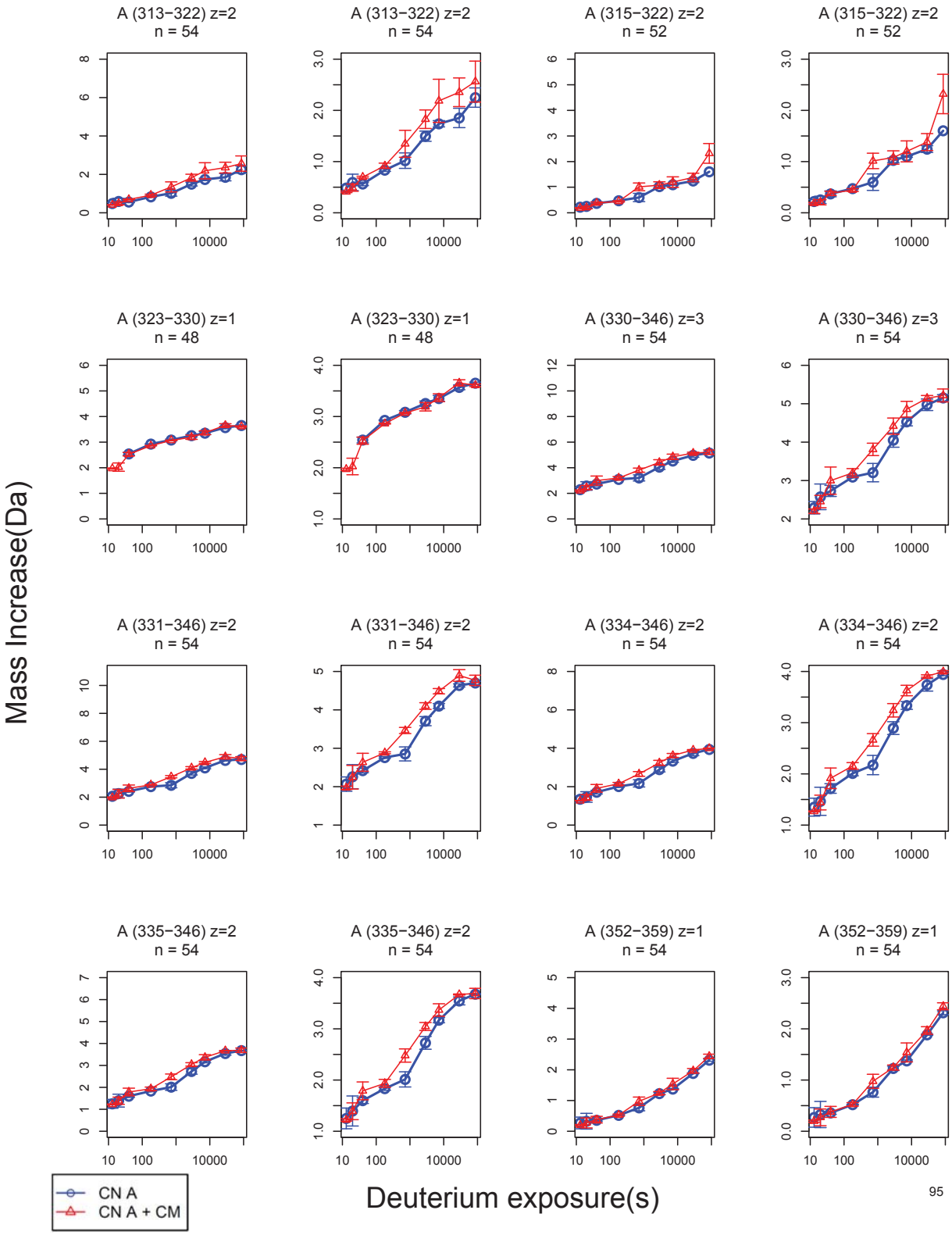

A (134−143) z=1
n = 44
A (134−143) z=1
n = 44
A (134−145) z=1
n = 54
A (134−145) z=1
n = 54
A (134−146) z=3
n = 54
A (134−146) z=3
n = 54
A (136−145) z=1
n = 54
A (136−145) z=1
n = 54
A (136−146) z=2
n = 54
A (136−146) z=2
n = 54
A (146−159) z=3
n = 54
A (146−159) z=3
n = 54
A (147−159) z=2
n = 52
A (147−159) z=2
n = 52
A (160−162) z=1
n = 52
A (160−162) z=1
n = 52
10
100
10000
Mass Increase(Da)
Deuterium exposure(s)
CN A
CN A + CM

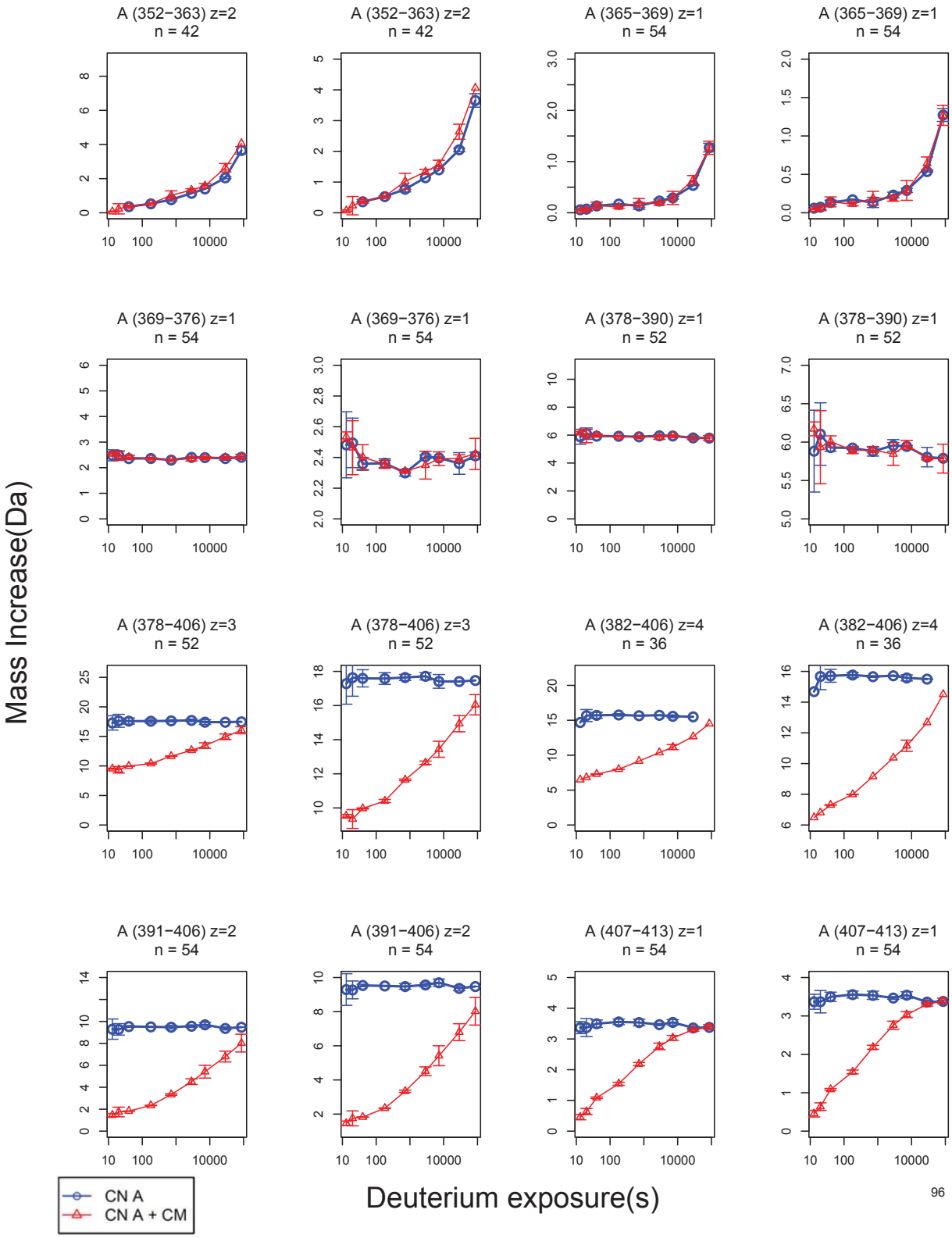

A (160−172) z=2
n = 46
A (160−172) z=2
n = 46
A (160−176) z=4
n = 52
A (160−176) z=4
n = 52
A (175−179) z=1
n = 43
A (175−179) z=1
n = 43
A (182−190) z=1
n = 54
A (182−190) z=1
n = 54
A (183−190) z=1
n = 54
A (183−190) z=1
n = 54
A (191−195) z=1
n = 42
A (191−195) z=1
n = 42
A (196−209) z=2
n = 52
A (196−209) z=2
n = 52
A (196−211) z=2
n = 54
A (196−211) z=2
n = 54
Mass Increase(Da)
Deuterium exposure(s)
CN A
CN A + CM

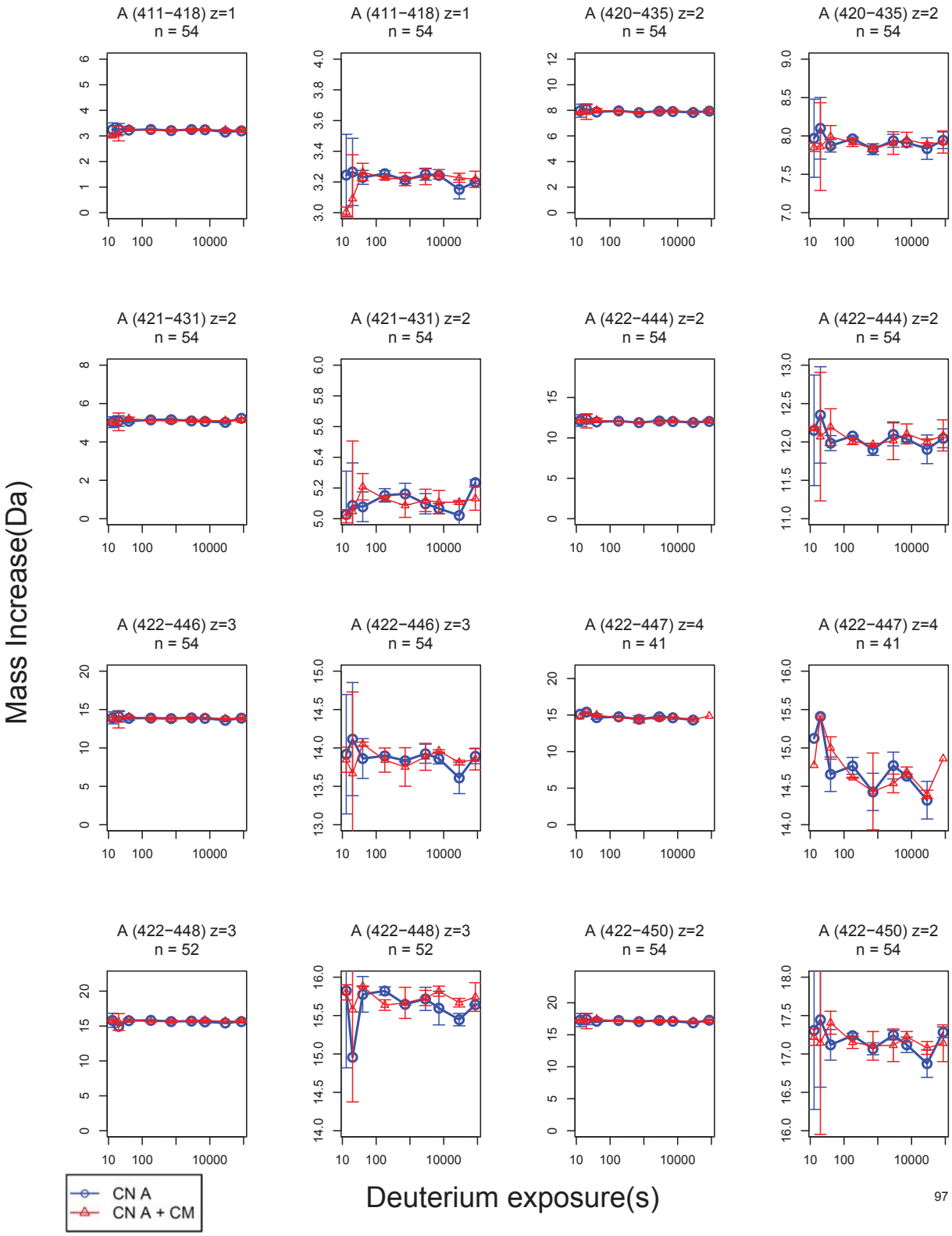

A (197−209) z=2
n = 54
A (197−209) z=2
n = 54
A (197−211) z=1
n = 52
A (197−211) z=1
n = 52
A (211−228) z=3
n = 54
A (211−228) z=3
n = 54
A (230−238) z=1
n = 54
A (230−238) z=1
n = 54
A (230−258) z=5
n = 54
A (230−258) z=5
n = 54
A (232−238) z=1
n = 54
A (232−238) z=1
n = 54
A (232−242) z=1
n = 54
A (232−242) z=1
n = 54
A (232−246) z=2
n = 50
A (232−246) z=2
n = 50
Mass Increase(Da)
Deuterium exposure(s)
CN A
CN A + CM

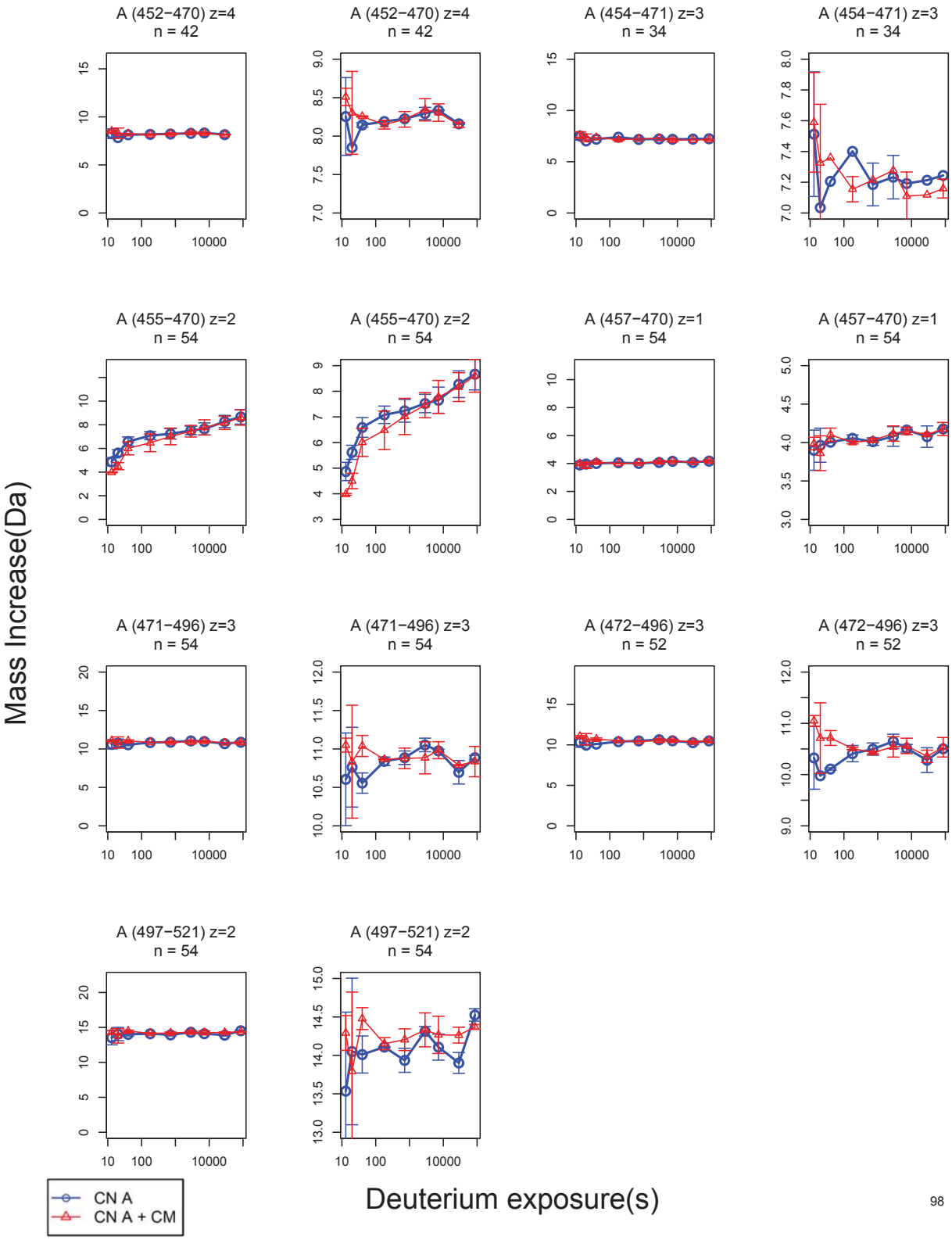

A (232−258) z=5
n = 52
A (232−258) z=5
n = 52
A (232−259) z=5
n = 54
A (232−259) z=5
n = 54
A (239−258) z=2
n = 54
A (239−258) z=2
n = 54
A (239−259) z=2
n = 54
A (239−259) z=2
n = 54
A (243−258) z=2
n = 26
A (243−258) z=2
n = 26
A (263−267) z=1
n = 52
A (263−267) z=1
n = 52
A (268−275) z=1
n = 54
A (268−275) z=1
n = 54
A (269−274) z=2
n = 50
A (269−274) z=2
n = 50
10
100
10000
Mass Increase(Da)
Deuterium exposure(s)
CN A
CN A + CM

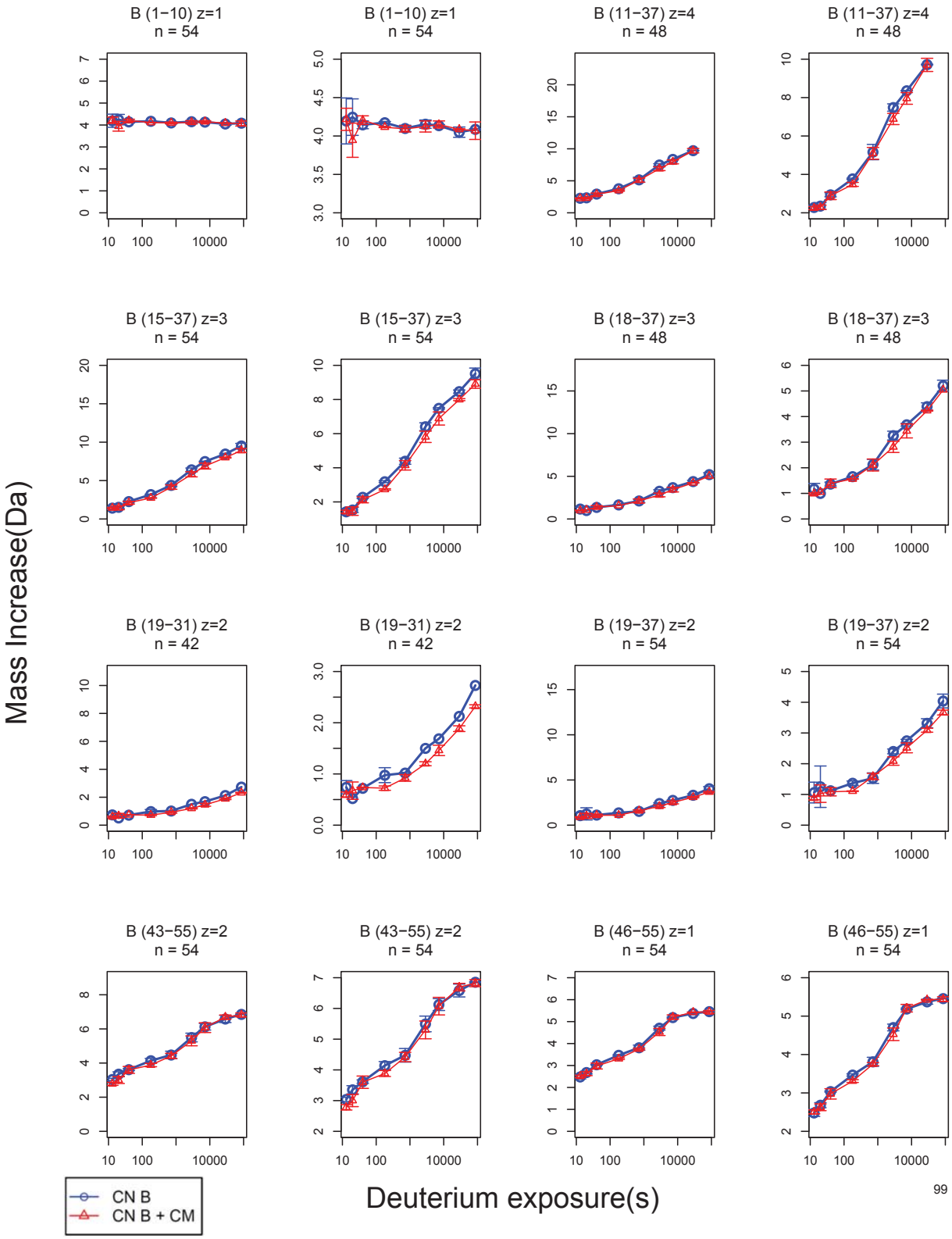

A (270−279) z=2
n = 54
A (270−279) z=2
n = 54
A (279−290) z=3
n = 54
A (279−290) z=3
n = 54
A (291−299) z=1
n = 54
A (291−299) z=1
n = 54
A (291−302) z=1
n = 54
A (291−302) z=1
n = 54
A (303−311) z=1
n = 54
A (303−311) z=1
n = 54
A (303−312) z=1
n = 54
A (303−312) z=1
n = 54
A (303−314) z=1
n = 54
A (303−314) z=1
n = 54
A (307−312) z=1
n = 44
A (307−312) z=1
n = 44
Mass Increase(Da)
Deuterium exposure(s)
CN A
CN A + CM

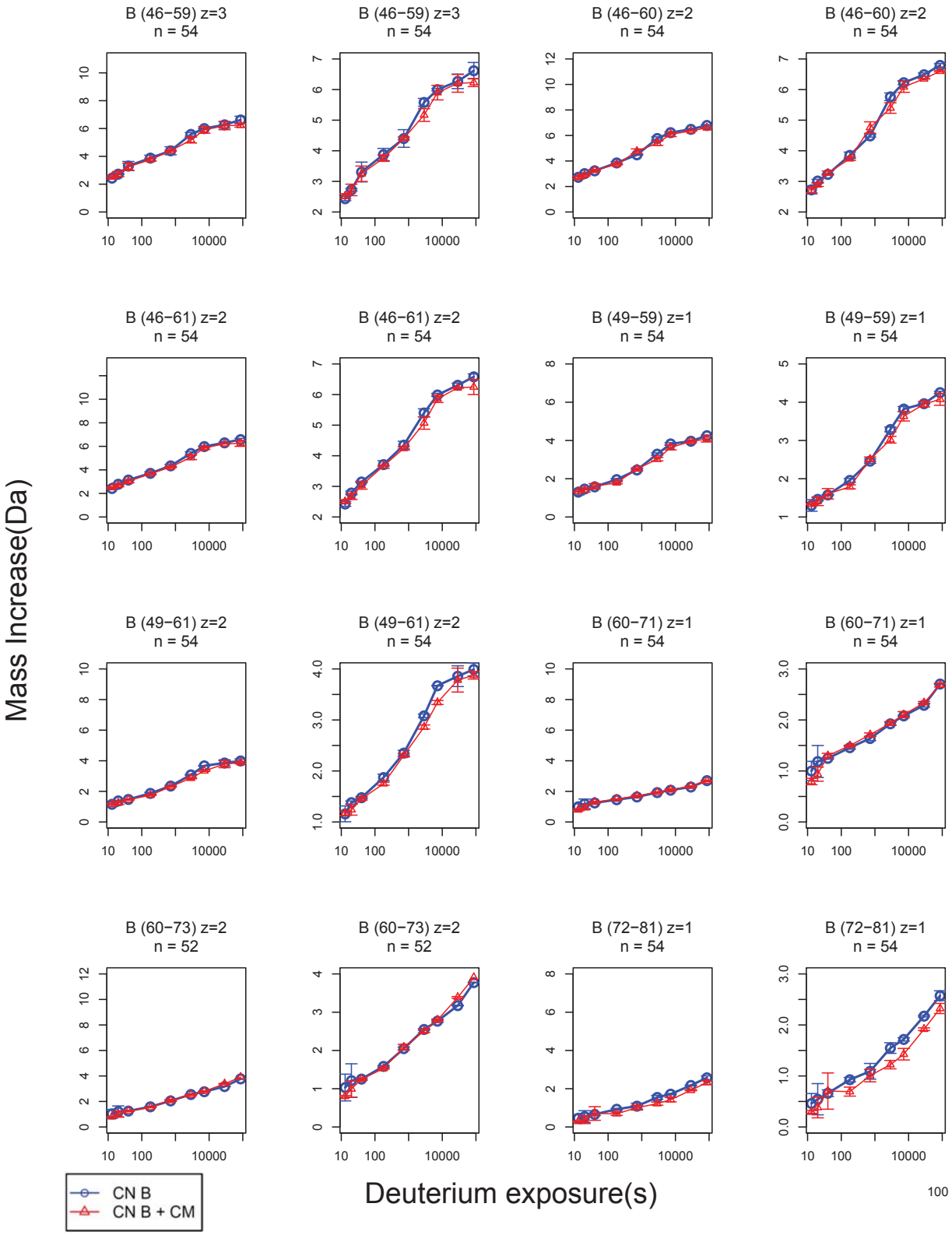

A (313−322) z=2
n = 54
A (313−322) z=2
n = 54
A (315−322) z=2
n = 52
A (315−322) z=2
n = 52
A (323−330) z=1
n = 48
A (323−330) z=1
n = 48
A (330−346) z=3
n = 54
A (330−346) z=3
n = 54
A (331−346) z=2
n = 54
A (331−346) z=2
n = 54
A (334−346) z=2
n = 54
A (334−346) z=2
n = 54
A (335−346) z=2
n = 54
A (335−346) z=2
n = 54
A (352−359) z=1
n = 54
A (352−359) z=1
n = 54
Mass Increase(Da)
Deuterium exposure(s)
CN A
CN A + CM

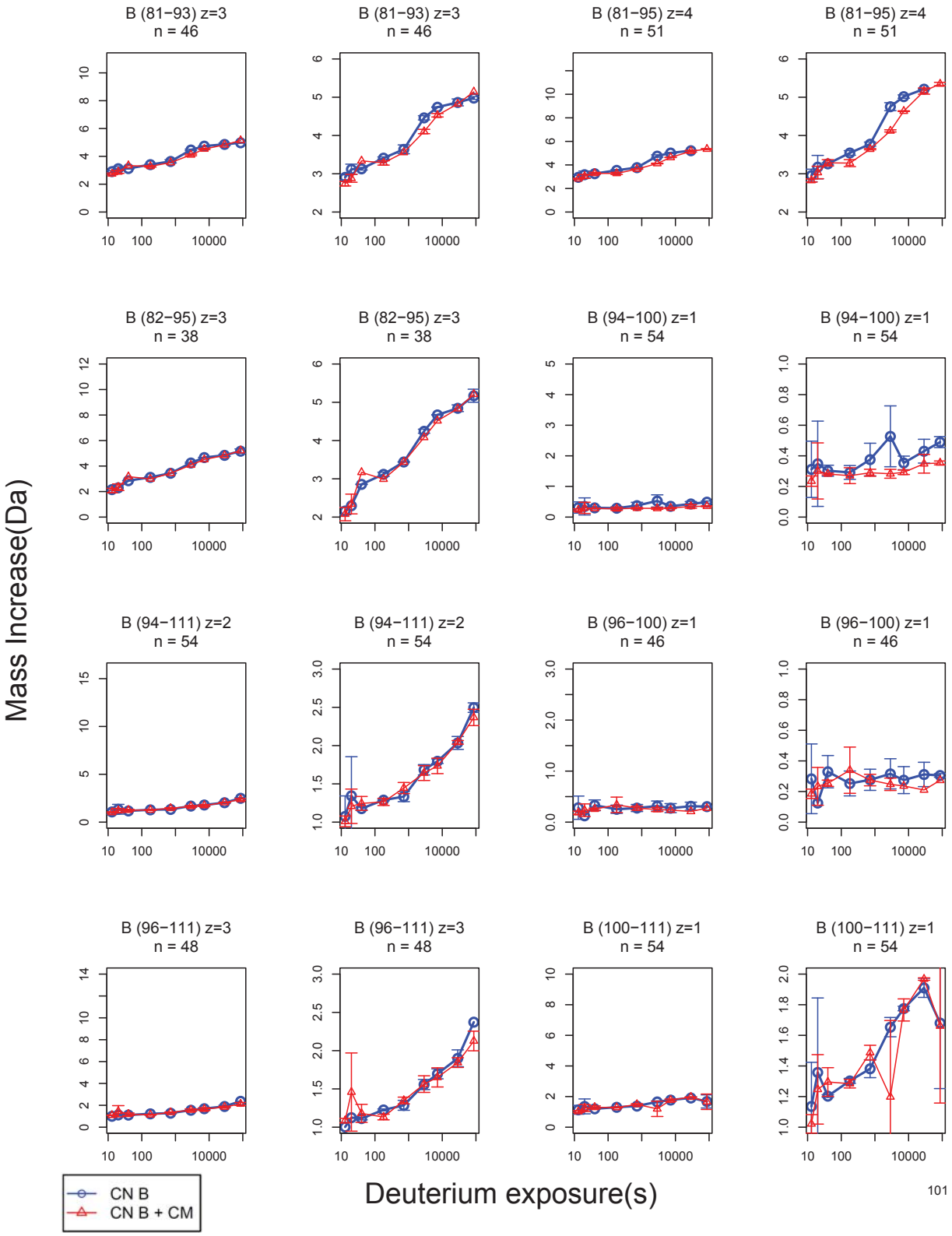

A (352−363) z=2
n = 42
A (352−363) z=2
n = 42
A (365−369) z=1
n = 54
A (365−369) z=1
n = 54
A (369−376) z=1
n = 54
A (369−376) z=1
n = 54
A (378−390) z=1
n = 52
A (378−390) z=1
n = 52
A (378−406) z=3
n = 52
A (378−406) z=3
n = 52
A (382−406) z=4
n = 36
A (382−406) z=4
n = 36
A (391−406) z=2
n = 54
A (391−406) z=2
n = 54
A (407−413) z=1
n = 54
A (407−413) z=1
n = 54
Mass Increase(Da)
Deuterium exposure(s)
CN A
CN A + CM

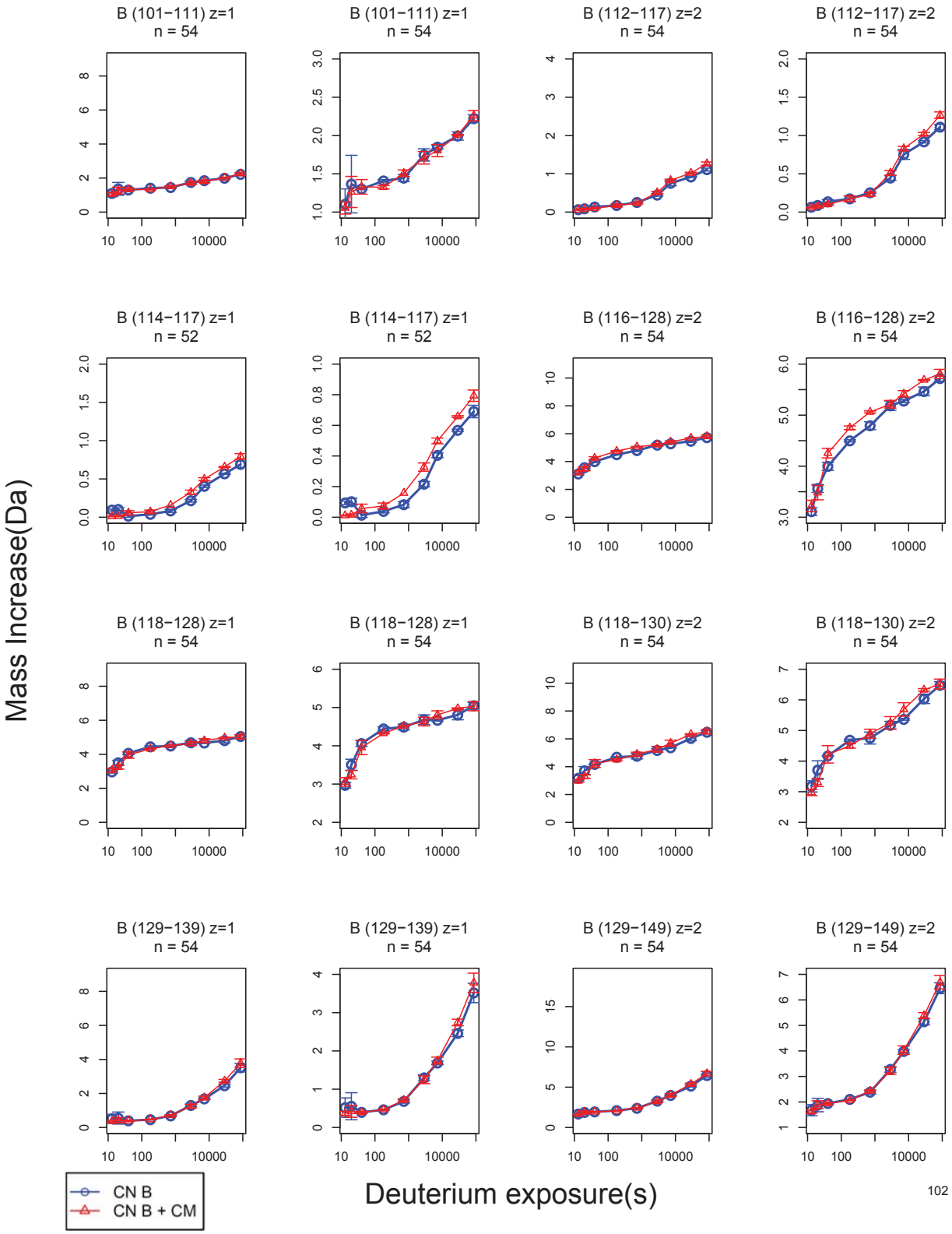

A (411−418) z=1
n = 54
A (411−418) z=1
n = 54
A (420−435) z=2
n = 54
A (420−435) z=2
n = 54
A (421−431) z=2
n = 54
A (421−431) z=2
n = 54
A (422−444) z=2
n = 54
A (422−444) z=2
n = 54
A (422−446) z=3
n = 54
A (422−446) z=3
n = 54
A (422−447) z=4
n = 41
A (422−447) z=4
n = 41
A (422−448) z=3
n = 52
A (422−448) z=3
n = 52
A (422−450) z=2
n = 54
A (422−450) z=2
n = 54
Mass Increase(Da)
Deuterium exposure(s)
CN A
CN A + CM


A (452−470) z=4
n = 42
A (452−470) z=4
n = 42
A (454−471) z=3
n = 34
A (454−471) z=3
n = 34
A (455−470) z=2
n = 54
A (455−470) z=2
n = 54
A (457−470) z=1
n = 54
A (457−470) z=1
n = 54
A (471−496) z=3
n = 54
A (471−496) z=3
n = 54
A (472−496) z=3
n = 52
A (472−496) z=3
n = 52
A (497−521) z=2
n = 54
A (497−521) z=2
n = 54
Mass Increase(Da)
Deuterium exposure(s)
CN A
CN A + CM


B (1−10) z=1
n = 54
B (1−10) z=1
n = 54
B (11−37) z=4
n = 48
B (11−37) z=4
n = 48
B (15−37) z=3
n = 54
B (15−37) z=3
n = 54
B (18−37) z=3
n = 48
B (18−37) z=3
n = 48
B (19−31) z=2
n = 42
B (19−31) z=2
n = 42
B (19−37) z=2
n = 54
B (19−37) z=2
n = 54
B (43−55) z=2
n = 54
B (43−55) z=2
n = 54
B (46−55) z=1
n = 54
B (46−55) z=1
n = 54
10
100
10000
Mass Increase(Da)
Deuterium exposure(s)
CN B
CN B + CM


B (46−59) z=3
n = 54
B (46−59) z=3
n = 54
B (46−60) z=2
n = 54
B (46−60) z=2
n = 54
B (46−61) z=2
n = 54
B (46−61) z=2
n = 54
B (49−59) z=1
n = 54
B (49−59) z=1
n = 54
B (49−61) z=2
n = 54
B (49−61) z=2
n = 54
B (60−71) z=1
n = 54
B (60−71) z=1
n = 54
B (60−73) z=2
n = 52
B (60−73) z=2
n = 52
B (72−81) z=1
n = 54
B (72−81) z=1
n = 54
Mass Increase(Da)
Deuterium exposure(s)
CN B
CN B + CM


B (81−93) z=3
n = 46
B (81−93) z=3
n = 46
B (81−95) z=4
n = 51
B (81−95) z=4
n = 51
B (82−95) z=3
n = 38
B (82−95) z=3
n = 38
B (94−100) z=1
n = 54
B (94−100) z=1
n = 54
B (94−111) z=2
n = 54
B (94−111) z=2
n = 54
B (96−100) z=1
n = 46
B (96−100) z=1
n = 46
B (96−111) z=3
n = 48
B (96−111) z=3
n = 48
B (100−111) z=1
n = 54
B (100−111) z=1
n = 54
Mass Increase(Da)
Deuterium exposure(s)
CN B
CN B + CM


B (101−111) z=1
n = 54
B (101−111) z=1
n = 54
B (112−117) z=2
n = 54
B (112−117) z=2
n = 54
B (114−117) z=1
n = 52
B (114−117) z=1
n = 52
B (116−128) z=2
n = 54
B (116−128) z=2
n = 54
B (118−128) z=1
n = 54
B (118−128) z=1
n = 54
B (118−130) z=2
n = 54
B (118−130) z=2
n = 54
B (129−139) z=1
n = 54
B (129−139) z=1
n = 54
B (129−149) z=2
n = 54
B (129−149) z=2
n = 54
10
100
10000
Mass Increase(Da)
Deuterium exposure(s)
CN B
CN B + CM

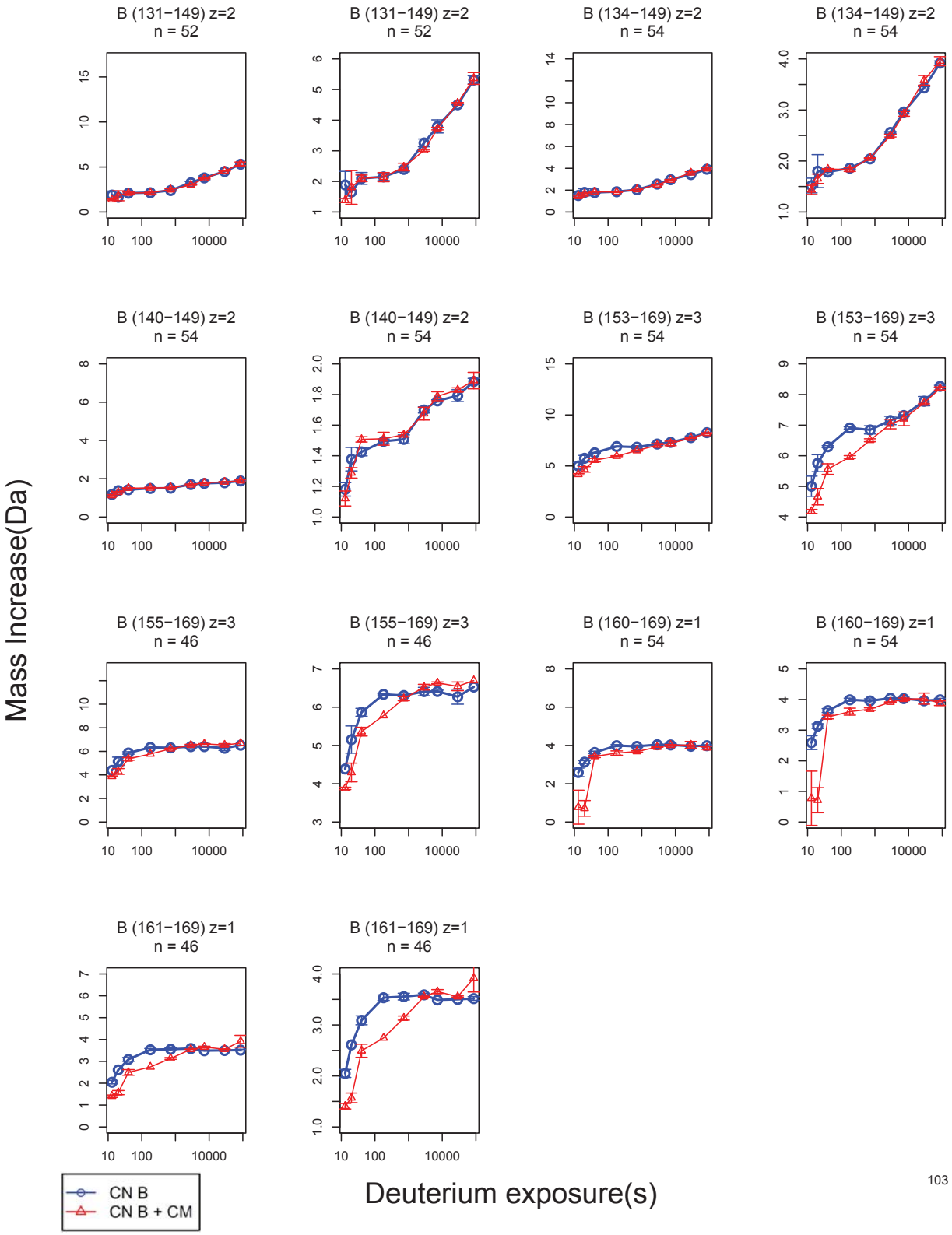

B (131−149) z=2
n = 52
B (131−149) z=2
n = 52
B (134−149) z=2
n = 54
B (134−149) z=2
n = 54
B (140−149) z=2
n = 54
B (140−149) z=2
n = 54
B (153−169) z=3
n = 54
B (153−169) z=3
n = 54
B (155−169) z=3
n = 46
B (155−169) z=3
n = 46
B (160−169) z=1
n = 54
B (160−169) z=1
n = 54
B (161−169) z=1
n = 46
B (161−169) z=1
n = 46
Mass Increase(Da)
Deuterium exposure(s)
CN B
CN B + CM

Figure 2.S2. Hydrogen exchange by peptides from CaN as measured by HX-MS comparing free CaN and bound CaN to the calmodulin. The identity of the peptides is shown at the top of each uptake plot. The blue lines represent uptake by the peptides of free CaN state and the red lines represent uptake by the peptides of bound CaN state.

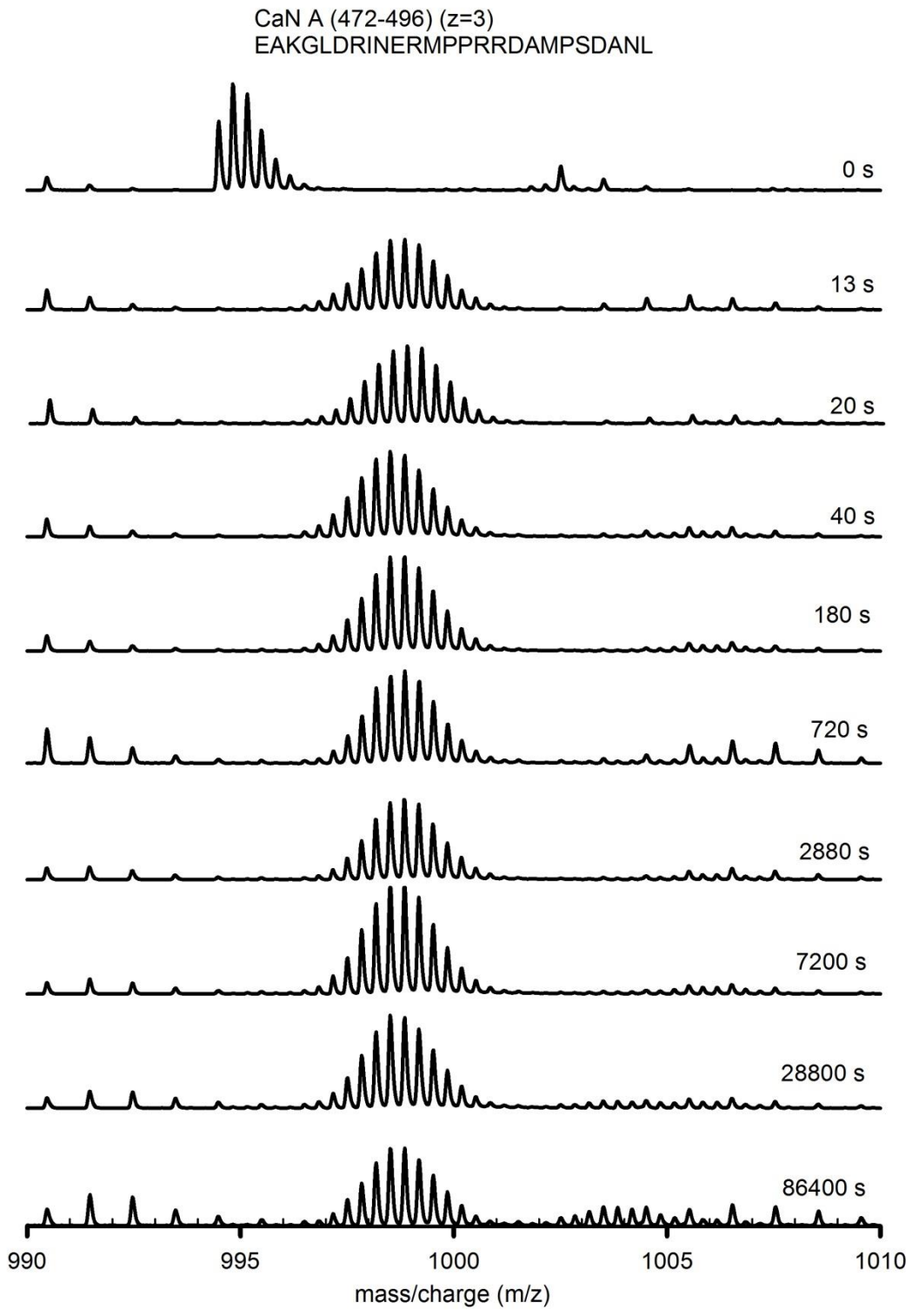


Figure 2.S3. Mass spectra of a representative peptide located in the AID region of calcineurin A in the free state at different labeling time points. The identity, charge state, and sequence are provided at the top of the spectra. The spectra confirm that there is no evidence of EX1 kinetics for the AID region of the CaN A.

Chapter Three

Exploring the best reference state to quantify the degree of structure in disordered proteins using millisecond HX-MS

3.1 Introduction

Intrinsically disordered proteins (IDPs) have regions that are highly flexible and lack stable secondary or tertiary structure.[1-6] Recently, there has been growing interest in IDPs due to their important role in many biological processes and functions. The biological functions include transcriptional regulation, cellular signaling, post-translational modification, and cellular recognition.[7-9] IDPs are associated principally with functions that require protein-protein interactions such as in cellular signaling.[7] Although many studies have shown that IDPs participate in functional interactions, less is known about the structural details of the interactions.

One of the key advantages of IDPs is their highly dynamic conformations that mediate protein-protein interaction through the phenomenon of coupled folding and binding.[10-12] In this type of interaction, folding of a disordered protein occurs upon engagement with its binding partner. Two modes of binding have been proposed. Conformational selection is one of the modes of binding proposed in IDP interactions. According to this mode of binding, transiently folded states, which exist within many conformational states of disordered proteins, can be selected by the binding partner to allow binding. The other mode of binding is the induced folding mode. In this case, the conformation flexibility of a single disordered protein can ease the interaction with its binding partner. This interaction leads to folding the disordered protein.[13, 14] More detailed information is required to understand the complexity of mechanisms of protein-protein interaction involving disordered proteins. Also, localized information about the presence or absence of transiently folded conformers of IDPs in the free state can be very important to provide support for one proposed mode of binding over the other.

Many techniques have been employed to provide information about structure of IDPs.[14-19] NMR and amide hydrogen exchange (HX) mass spectrometry are among the techniques that

can measure the degree of protein structure with high spatial resolution in IDPs. The rate of amide HX is highly dependent on hydrogen bonding and solvent accessibility which is related to the conformational flexibility of the protein backbone.[20, 21] Highly flexible regions will exchange more rapidly, while rigid regions exchange more slowly.[22] Millisecond HX can be a very useful technique to characterize IDPs because most of the exchange occurs on this timescale at room temperature and near neutral pH.[23-26] Protection in HX experiments, which is a measure of the slowing of exchange caused by structure, can reveal information about protein structures. The degree of protection can be directly related to the degree of structure if the observed rate of HX is compared to the rate of HX by an unprotected reference state. The unprotected reference state should represent the fastest possible exchange in order to quantify the degree of protection. A higher degree of structure causes the more protection. Without an unprotected reference state, it is impossible to estimate such protection because the unprotected reference state is an essential part in measuring protection. Generally, the chemical exchange, calculated using an empirical formula based on unstructured model peptides[27], is used as an unprotected reference state. Using chemical exchange as a reference state here is the key in measuring the protection. Any inaccuracy in the unprotected reference state would underestimate or overestimate the degree of structure calculation. Our lab[24] and others[28, 29] have found that the calculated chemical exchange does not seem to be accurate enough to be used as an unprotected reference state in weakly protected IDPs. In some cases, the chemical exchange calculations were slower than HX by unstructured peptides.

In this work, we explore different approaches to prepare an unprotected reference state for quantifying the degree of structure in IDPs. We employed a quench-flow device, built in-house,[30] for HX labeling on the millisecond timescale. The interaction domains of the activator

for thyroid and retinoid receptors (ACTR) and the CREB binding protein (CBP) were used as model IDPs to explore the best approach to produce an unprotected reference state for millisecond HX. ACTR is a near-random coil IDP that has some residual helicity[11, 23, 31, 32] while CBP is a molten globular IDP that transiently becomes unstructured as revealed by NMR and HX-MS.[23, 24, 32, 33] We have tested three different methods that can be used to prepare an unprotected reference state. These methods are chemical exchange calculations, addition of denaturing agents, and millisecond HX labeling of peptic peptides obtained from the IDP. In addition, two unstructured model peptides have been used as internal standards to see how changes in solution could alter chemical exchange. We have found that the peptic reference peptide can be used as an accurate unprotected reference state for determining the degree of the structure.

3.2 Materials and methods

3.2.1 Protein expression and purification

ACTR$^{1023\text{-}1093}$ (UniProt NCOA3_HUMAN) and CBP$^{2059\text{-}2117}$ (UniProt CBP_MOUSE) domains were co-expressed from a pET22B co-expression vector[11] in *Escherichia coli* BL21 (DE3) and separately purified.[23] A single colony was selected to inoculate 25 mL of LB medium after overnight incubation at 37 °C on ampicillin/LB agar plates. The cells were incubated overnight at 37 °C in a rotary shaker at 225 rpm. 10 ml of the starter culture was then transferred to 500 mL of LB medium containing ampicillin. The cultures were induced with 1 mM IPTG after the OD_{600} had reached 1.0 and incubated for an additional 4 h at 37°C in a rotary shaker at 225 rpm. The cells were then pelleted at 5000g for 10 min at 4°C and then stored at –80 °C. Ten mL of lysis buffer [20 mM Tris, 1 mM EDTA, 1g/L lysozyme, and 5 units benzoase nuclease (pH 7.5)] with (250 µL/g pellet) protease inhibitor cocktail [23 mM AEBSF, 2 mM

bestatin, 0.3 mM pepstatin A2, and 0.3 mM E-64] were used to resuspend the cell pellet by gentle agitation at room temperature for 10 min. The lysate were centrifuged at 15000g for 10 min at 4°C to remove cell debris. A fast protein liquid chromatography FPLC system (AKTAPrime Plus, GE Healthcare, Piscataway, NJ) was used to purify proteins at 4 °C using 214 nm detection. Ion exchange was used to isolate ACTR and CBP as they have different predicted pI values of 4.2 and 11.1, respectively. For ACTR purification, the lysis buffer was first exchanged for anion exchange buffer (50 mM piperazine pH 9.0) using a desalting column (HiPrep 26/10 Desalting, GE Healthcare). Protein fractions containing ACTR were pooled after identification by SDS-PAGE and passed through an anion exchange column (HiTrap Q XL, GE Healthcare) in 50 mM piperazine (pH 9.0) using a 0 to 1 M NaCl gradient over 70 min at a rate 1 mL/min. After that, fractions containing ACTR were pooled and passed through a gel filtration column (HiPrep 16/60 Sephacryl S-100 HR, GE Healthcare) in 10 mM sodium phosphate and 50 mM NaCl buffer pH 6.9. For CBP purification, the lysis buffer was first exchanged for cation exchange buffer (50 mM boric acid pH 9.0) containing 3 mM dithiothreitol, to avoid methionine oxidation, using a desalting column (HiPrep 26/10 Desalting, GE Healthcare). Protein fractions containing CBP were pooled after identification by SDS-PAGE and passed through a cation exchange column (HiTrap SP FF, GE Healthcare) in 50 mM boric acid and 3 mM DTT buffer (pH 9.0) using a 0 to 1 M NaCl gradient over 70 min at a rate 1 mL/min. After that, fractions containing CBP were pooled and passed through a desalting column in 10 mM sodium phosphate and 50 mM NaCl buffer pH 6.9. Protein stocks were split into aliquots and frozen and stored at –80°C. An aliquot was then taken and dialyzed against 20 mM sodium citrate and 100 mM NaCl buffer (pH 6.5) for 2 hours followed by overnight dialysis at 4 °C using mini dialysis device (Slide-A-Lyzer 2000 MWCO, Thermo Scientific). The ACTR and CBP concentration were then

determined to be 18.8 µM and 17.2 µM, respectively, using a BCA assay[34] standardized with BSA (Thermo Fisher Scientific).

3.2.2 Preparation of peptic reference peptides

Before starting the preparation of peptic peptides, ACTR was concentrated by freezing 3 mL of stock with liquid nitrogen followed by lyophilizing using LABCONCO Freezone (model 7670521, Kansas City, MO, USA). The ACTR was reconstituted in 1 mL of ultrapure water. Then the sample was dialyzed against citrate buffer (pH 6.5) two times for two hours each followed by overnight dialysis at 4 °C using mini dialysis device (Slide-A-Lyzer 2000 MWCO, Thermo Scientific). The concentration of ACTR stock was 45.2 µM. ACTR peptic peptides were prepared by diluting 50 µL of (45.2 µM) ACTR to 2.2 mL with 0.1% formic acid and 150 µL of stock CBP (17.2 µM) was diluted to 2.4 mL with 0.1% formic acid. The samples were passed through an immobilized pepsin column, described in the following section, using a syringe pump at a flow rate 50 µL/min at room temperature. The digested sample was dried using a LABCONO CentriVap Concentrator (model 780016) combined with Freezone for 1 hour at 40 °C. The dried samples were reconstituted in 50 µL of citrate buffer (pH 6.5), frozen, and stored at –80 °C.

3.2.3 Hydrogen exchange labeling

Millisecond hydrogen exchange (HX) was conducted using a home-built quench-flow device as described previously[24] (see section 1.3.3.3) with some improvements in syringe types and fittings. ChemSeal syringes with 1/4-28 thread tip (Hamilton, Reno, NV) were used instead of syringes with luer lock tip to prevent leaks at high flow rate because this type of syringe has a pressure limit of 14 bar. All three syringes were connected to three lines of fused silica capillaries with adapter assembly (P-627 female 10-32 coned to female 1/4-28 thread

configuration, IDEX, Oak Harbor, WA). NanoTight tubing sleeves (F-242) and fittings (F-300) (IDEX, Oak Harbor, WA) for the assembly were used to connect the silica capillaries and the assembly. All other parts of the quench-flow system remained the same as described previously in more detail.[24] ACTR and CBP were labeled with separate sets of experiments using the quench-flow device. The ACTR (8 µM) and CBP (8 µM) proteins were compared with their reference peptides ACTR (3 µM) and CBP (3 µM), respectively. All four samples were spiked with two internal standard peptides FKPGI (GenScript, Piscataway, NJ) and YPI (AnaSpec, Fremont, CA) used here as rapidly and slowly exchanging peptides, respectively. The samples, in H_2O buffer (20 mM sodium citrate, 100 mM NaCl, pH 6.5), were labeled with a 5-fold excess of D_2O buffer (20 mM sodium citrate, 100 mM NaCl, pD 6.5). According to the applied flow rates, the labeling time ranged between 42 and 3500 milliseconds. The labeled samples were then quenched at a 5:6 volume ratio with quench buffer (200 mM sodium phosphate, pH 2.3). Then the quenched samples were immediately flash-frozen with liquid nitrogen and kept at –80°C until analysis. Fully deuterated controls were labeled manually with a 5-fold excess of D_2O buffer for 18 hours and then quenched at a 5:6 volume ratio with the quenching buffer. Undeuterated controls, the samples were diluted 5-fold with H_2O buffer instead of D_2O buffer and followed the steps for quenching and flash-freezing the samples. All HX labeling was carried out at room temperature (23±1°C) in triplicate for each labeling time point.

To measure chemical exchange in different buffers, a slowly exchanging model peptide (YPI) was labeled for between 20 to 4000 sec (standard timescale) using an H/D-X PAL robot (LEAP Technologies, Carrboro, NC). Three deuterium labeling buffers, (20 mM sodium citrate, 100 mM NaCl, pD 6.5, control), (3 M deuterated guanidine hydrochloride with 97%D, 20 mM sodium citrate, 100 mM NaCl, pD 6.5) and (6 M Urea-d6 (99.9%), 20 mM sodium citrate, 100

mM NaCl, pD 6.5), were used to label the sample at 5:1 (v:v) ratio. The exchange was quenched by 5:6 dilution in a quench buffer (200 mM sodium phosphate, pH 2.3). The quenched samples were analyzed immediately. Fully deuterated guanidine hydrochloride was initially prepared by dissolving 1.5 g of guanidine hydrochloride in 15 ml D_2O (99.9%D). The solution was left overnight for complete deuteration at room temperate. Then the solution was flash-frozen followed by an overnight freeze-drying cycle. A second round of dissolution/lyophilization of guanidine in D_2O was repeated to obtain more deuteration. The degree of deuteration of guanidine hydrochloride, as measured by ^1H NMR on a Bruker DRX-500 spectrometer, was 97%D (see Figure 3.S1).

3.2.4 Chemical exchange calculations

The rate of chemical exchange by amide hydrogens was calculated using a sum of exponentials for each amide in a peptide using equation (1).[27]

$$\Delta m\,(t) = \sum_{i=3}^{n} [1 - \mathrm{e}^{-k_{\mathrm{ch}}^{(i)} t}] \qquad (1)$$

Where n and $k_{\mathrm{ch}}^{(i)}$represent the number of residues and the rate constant of chemical exchange[27] for residue *i* of a peptide, respectively. Δm is the extent of deuteration at D_2O exposure time *t*. As it shown in the equation (1), the value of *i* begins with 3. The reason is that the back-exchange is quite fast for the first two residues of the peptide during proteolysis and LC steps.[27, 35] The rate constant for proline residues was zero because there is no amide hydrogen. To calculate chemical exchange, we used an Excel spreadsheet which calculates chemical HX based on empirically derived formulas[27, 35] adapted from a spreadsheet provided by the Englander Lab (available online http://hx2.med.upenn.edu/). To compare the HX kinetics in experimental and theoretical states, we normalized the chemical exchange, $\Delta m_{\mathrm{ch}}(t)$, to the ratio between the

measured mass increase, Δm_{QF} , of the latest measured time point for each peptide, t_{max} and the calculated mass increase for chemical exchange, Δm_{ch}, at the same t_{max}.

$$\Delta m_{\mathrm{ch}}(t)(\mathrm{normalized}) = \Delta m_{\mathrm{ch}}(t) \times \frac{\Delta m_{\mathrm{QF}}(t_{\mathrm{max}})}{\Delta m_{\mathrm{ch}}(t_{\mathrm{max}})} \qquad (2)$$

To measure differences in the chemical exchange rate caused by different solution conditions, we measured hydrogen exchange by an unstructured peptide (YPI) on the seconds-to-hours timescale (standard timescale). The HX data were collected in 20 mM sodium citrate, 100 mM NaCl, pD 6.5 buffer without denaturants (control) and in the presence of 3 M guanidine or 6 M urea. Since YPI has only one amide that undergoes measurable exchange, the hydrogen exchange data for YPI peptide were fit to the equation (3) using non-linear least squares regression (SigmaPlot 12.5, Systat Software Inc., San Jose, CA) as shown below:

$$\Delta m\,(t) = y + a\,(1 - \mathrm{e}^{-k\,t}) \qquad (3)$$

Where *k* represents the rate constant for exchange and *a* represents the deuterium recovery. The best fits were achieved by incorporating *y* in the equation and *y* represents deuteration at *t* equal to zero. The best fits were also achieved for each condition within 10-20 iterations.

3.2.5 LC-MS analysis

ACTR peptic peptides were assigned using MS data from a time-of-flight mass spectrometer (Agilent 6220) and MS/MS data from a linear ion trap (Thermo LTQ-XL), as described previously.[(36)] CBP peptides were assigned using a quadrupole time-of-flight mass spectrometer (Agilent 6530). Several different gradients were used to separate and identify CBP peptides. MS/MS spectra of CBP peptides were assigned using Mass Hunter Qualitative Analysis (version B.06.00). Each MS/MS spectrum was examined to validate the assignment of

the CBP peptides and take out all potential in-source fragmented peptides that have same C-terminus with same retention time.

All LC-MS analysis of labeled sample was carried out on a time-of-flight mass spectrometer (Agilent 6220, Santa Clara, CA, USA) combined with HPLC (Agilent 1200 series). Individual frozen samples were thawed by hand for 3 minutes immediately before injection into a refrigerated column compartment built in-house.[30] The temperature of the compartment was maintained at 0°C. The LC mobile phases were 0.1% formic acid in water (A) and 90% acetonitrile/10% water with both containing 0.1% formic acid (B). After injection, the protein samples were digested online through an immobilized pepsin column[37, 38] (self-packed in a 2.1 mm × 100 mm column) at 200 µL/min with A. The peptides were trapped and desalted over 4 min on a reversed phase C12 trap (Jupiter 4 µm Proteo 90 Å, C12 (Phenomenex) packed by Optimize Technologies (Apple Valley, Minnesota) in a 1.5 × 5 mm OPTI-LYNX II cartridges). The resulting peptides were then separated on a C18 column (ZORBAX 300SB, 3.5 µm, 1 × 50 mm) using a gradient (5-60% B over 6 min for ACTR peptides, 5-45% over 6 min for CBP peptides) at 50 µL/min. All mass spectra were measured in positive ESI mode with a capillary voltage of 4000 V, drying gas flow 10 L/min, fragmentor 150 V, and temperature of 325°C. HDExaminer (version 2.0, Sierra Analytics, Modesto, CA) was used for initial deuterium calculation and data analysis. After that, the time required to reach 50% deuteration ($t_{50\%}$) for each peptide in both states of proteins was measured.[36] For the peptide-level $t_{50\%}$ calculation, the deuterium uptake values were exported from HDExaminer to a spreadsheet and normalized to 100% using the totally deuterated controls. Then, the $t_{50\%}$ values for each state were calculated using linear interpolation between the two points that spanned 50% exchange. The $t_{50\%}$ ratio (hereafter referred to as protection ratio) was then determined from the $t_{50\%}$ value of the protein

segment measured under native condition divided by the $t_{50\%}$ value of the unprotected reference state. Since there were many overlapping peptides for ACTR but not for CBP, we calculated the residue-resolved protection ratio. The ACTR residue-resolved $t_{50\%}$ ratios were determined as explained in previous work.[36] We used the equation 4 to calculate the value, as follows:

$$t_{50\%,j} = \frac{\sum_i (E_{i,j}.t_{50\%,i}.w_i)}{\sum_i w_i} \qquad (4)$$

First, we mapped the protection ratio on each residue, *j*, of a given peptide, *i*, of the ACTR protein. Then, a separate exchangeability matrix $E_{i,j}$ was made for each residue in each peptide. The value of $E_{i,j}$ is one if the residue reports on the exchange, while the value is zero if the residue does not report on the exchange. All proline residues, which do not have amide hydrogens, and first two residues in a given peptide, which undergo rapid back-exchange under quench conditions[27, 35], have $E_{i,j}$ equal to zero. Equation 4 also takes the peptide length into consideration by using a weighting function, w_i, equal to $(\sum_i E_{i,j})^{-2}$. The weighting function is used to give more weight to shorter peptides with fewer exchangeable amide hydrogen rather than long peptides because shorter peptides give better spatial resolution. The residue-resolved $t_{50\%}$ ratio was then calculated from the given equation values.

3.3 Results

Hydrogen exchange (HX) can provide structural details about intrinsically disordered proteins (IDPs) with high spatial resolution. Since the rate of hydrogen exchange is highly dependent on the conformational flexibility of proteins, most of HX occurs on the millisecond timescale at neutral pH and room temperature for IDPs. With a quench-flow apparatus, we can perform HX at timescales ranging from 42 – 3500 msec. We used ACTR and CBP as model systems for conformational analysis of IDPs. ACTR is a mostly unstructured protein that

transiently occupies a folded state as described by NMR and HX-MS.[11, 23, 24, 31, 32] In contrast, CBP is a mostly helical molten globule that transiently occupies an unfolded state as described by NMR and HX-MS.[23, 39-41] In order to determine the degree of structure in IDPs by HX, an accurate unprotected reference state is required to avoid any misleading information about the structure. The unprotected reference state here represents HX by the protein in its fully unprotected state giving the fastest possible rate of HX. Slower exchange by a segment compared to its unprotected reference state indicates the region is structured.

In this chapter, we explore different approaches to prepare an unprotected reference state that can be used for mapping the degree of structure in IDPs. These approaches are chemical exchange calculation, addition of denaturing agents, and HX labeling of peptic peptides. In an ideal case, the rate of exchange by the reference will be the same as the exchange by an unstructured peptide. For chemical exchange, the unprotected reference state is based on theoretical calculations. We used free peptides to test whether the measured rate of exchange matched the calculated rate. For addition of denaturing agents, we tested with free peptides to see whether unstructured peptides have the same rate of HX in denaturants. Finally, for peptic reference peptides, we compared HX in the unstructured regions of the proteins with HX by the same regions as isolated peptides, and then compare measured protection to known secondary structural elements.

3.3.1 Calculated chemical exchange as an unprotected reference state

To evaluate the accuracy of calculated chemical exchange as an unprotected reference state, we compared HX data from unstructured peptides with calculated exchange. If chemical exchange calculations are accurate then we would expect that the measured rate of HX would be the same as calculated exchange. Two representative deuterium uptake plots of peptic reference

peptides from ACTR (residues 1023-1032) and CBP (residues 2059-2068) are shown in Figure 3.1. The white filled points denote the peptic reference peptide deuterium uptake from the proteins and the black filled triangles denote the calculated deuterium uptake for the corresponding peptides. As shown in Figure 3.1a and 3.1b, the reference peptides from ACTR and CBP exhibit faster exchange than predicted by calculation. There is a discrepancy between measured and predicted HX and this discrepancy is the case for all the reference peptides from ACTR and CBP (see Figure 3.S2). To determine if the difference in the exchange is not only related to ACTR and CBP peptides, we used two model peptides, YPI and FKPGI. Recently, an internal standard peptide (PPPI) was used in an HX experiment to compensate for any possible differences in chemical exchange rates caused by different solution conditions.[(29)] Here, we have used two unstructured model peptides, FKPGI and YPI, to compare the HX predicted by calculation with the HX by the peptides measured under native condition. Our model peptides were selected by selecting some characteristics of unstructured peptides. These two peptides are short which reduces the likelihood of folding and they contain proline which breaks a possible helix due to its side chain steric interferences. The residues of the two peptides also have poor helix-forming propensity.[(42)] All these criteria make the YPI and FKPGI useful models of unstructured polypeptides Because the first two amide hydrogens undergo a rapid back-exchange under quench conditions and proline does not have an amide hydrogen, YPI has only one slow exchangeable residue, isoleucine. YPI exchanges at standard timescale and can be described by first-order kinetics.[(43)] FKPGI has two amides that undergo measurable exchange, isoleucine and glycine, with slow and fast rates of exchange, respectively.[(27)] We compared the HX data of the model peptide (FKPGI) with the calculated exchange on the millisecond timescale. As shown in Figure 3.1c, again HX by the model peptide is faster than calculated exchange. To confirm that

the differences in the exchange are not related to our quench-flow system, we also used conventional labeling for YPI deuteration under native conditions at the standard labeling timescale and compared it with chemical exchange calculation. The results were the same as for FKPGI in millisecond timescale. The exchange by YPI was faster than exchange predicted by chemical exchange calculation as shown in Figure 3.1d. The rate of exchange differences of the chemical exchange with already unstructured peptides may yield unreliable calculation for the degree of structure of the IDPs. Therefore, the chemical exchange calculation may not be sufficiently accurate for an unprotected reference state for weakly protected and unprotected regions in proteins.

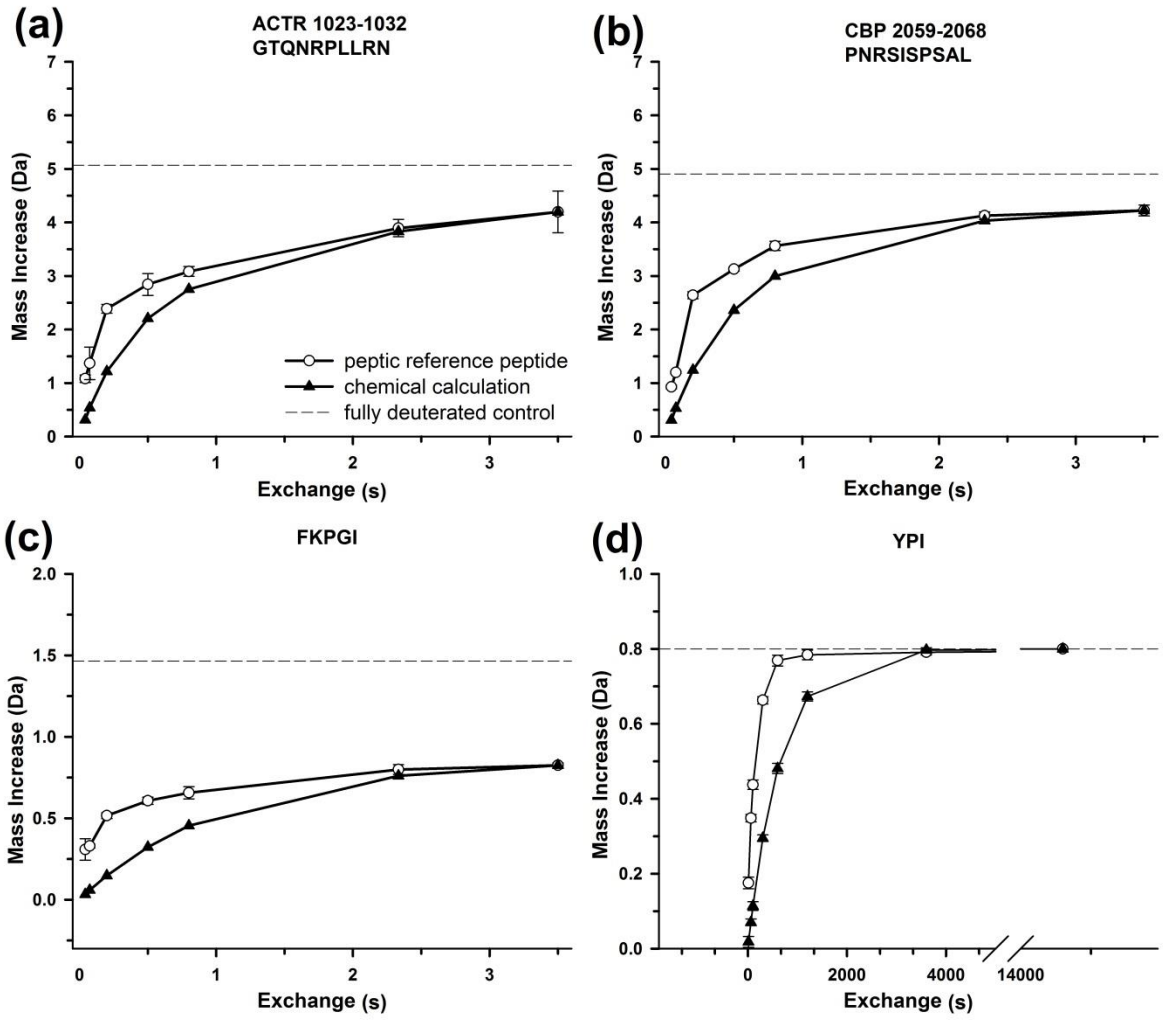


Figure 3.1. Deuterium uptake curves for representative peptides from (a) ACTR, (b) CBP, and (c and d) two unstructured model peptides. The open circles represent uptake by the peptic reference peptides and the black filled triangles represent the chemical exchange calculation normalized to match the latest measured time point of each respective curve. The error bars denote standard deviation from triplicate measurements. The horizontal axis in each plot is the exchange time in seconds and the vertical axis is the number of exchangeable amides by each peptide. The horizontal dashed line represents the maximum deuterium uptake measured after exchange for 12 h. All the data were acquired with the quench-flow system between 42 – 3500 milliseconds except for YPI (d) where the data were acquired at standard timescales. See text for additional details.

3.3.2 Protein HX in the presence of denaturants as an unprotected reference state

Another possibility is to prepare an unprotected reference state by incubating the proteins in denaturants before millisecond HX labeling. Measuring HX in the presence of denaturants, unlike chemical exchange calculation, is based on experimental measurement. The denaturing agents may promote unfolding of proteins both directly by disrupting hydrogen bonding and indirectly by altering water structure and dynamics mechanism[(44)] to make even disordered proteins completely unstructured. Denaturing conditions remove protection thereby causing the rate of exchange to become as fast as possible. For this purpose, we tested two common denaturing agents, urea and guanidine. We carried out the HX experiments with labeling time points ranging from 20 sec to 4000 sec. We compared the HX by the slowly-exchanging unstructured peptide, YPI, in the absence of denaturants (control) with the presence of 3 M guanidine or 6 M urea in citrate buffer (pD 6.5). YPI is unstructured, and expected to have no protection when HX data in the absence of denaturant is compared with the presence of denaturants. Thus, we would expect the rate of exchange by YPI in the absence and presence of denaturants would be the same if this method is a good unprotected reference state. The HX data of the model peptide in these conditions are shown in Figure 3.2. The black filled points denote deuterium uptake by the peptide under a non-denatured condition in citrate buffer (pH/pD 6.5), and the white filled triangles and squares denote deuterium uptake in the presence of guanidine (3 M) and urea (6 M), respectively. Since YPI has only one exchangeable residue and is well-described by single exponential function, each curve was fitted as described by equation 3 in the Materials and Methods section. The curve fitting parameters are listed in Table 3.1. From the curve fitting, rate constants for each state were calculated. The rate constant for YPI in the absence of denaturant, the presence of 3 M guanidine, and the presence of 6 M urea conditions

were 0.0058, 0.0085, and 0.0031 s^{-1}, respectively (Table 3.1 and Figure 3.2). Urea and guanidine have significant and opposing effects on hydrogen exchange. In the case of adding guanidine to the buffers, the model peptide exchanged more quickly than in the absence of denaturant. The rate of HX in guanidine-added conditions speeds up by a factor of 1.46. While in the case of adding urea to the buffers, there is an opposite effect. The model peptide exchanged more slowly in the urea-added conditions and the rate of HX in the presence of 6 M urea slows down by a factor of 1.87. Because YPI peptide is designed not to have any conformational protection, the fast exchange in guanidine and slow exchange in urea is a reflection of the differences in the solution conditions that alter the chemical exchange process. These differences may not affect the actual protection coming from the degree of structure of proteins. Therefore, the addition of denaturants to the proteins may not work as an unprotected reference state for IDPs because it would underestimate the degree of structure in the case of urea or overestimate the degree of structure in the case of guanidine.

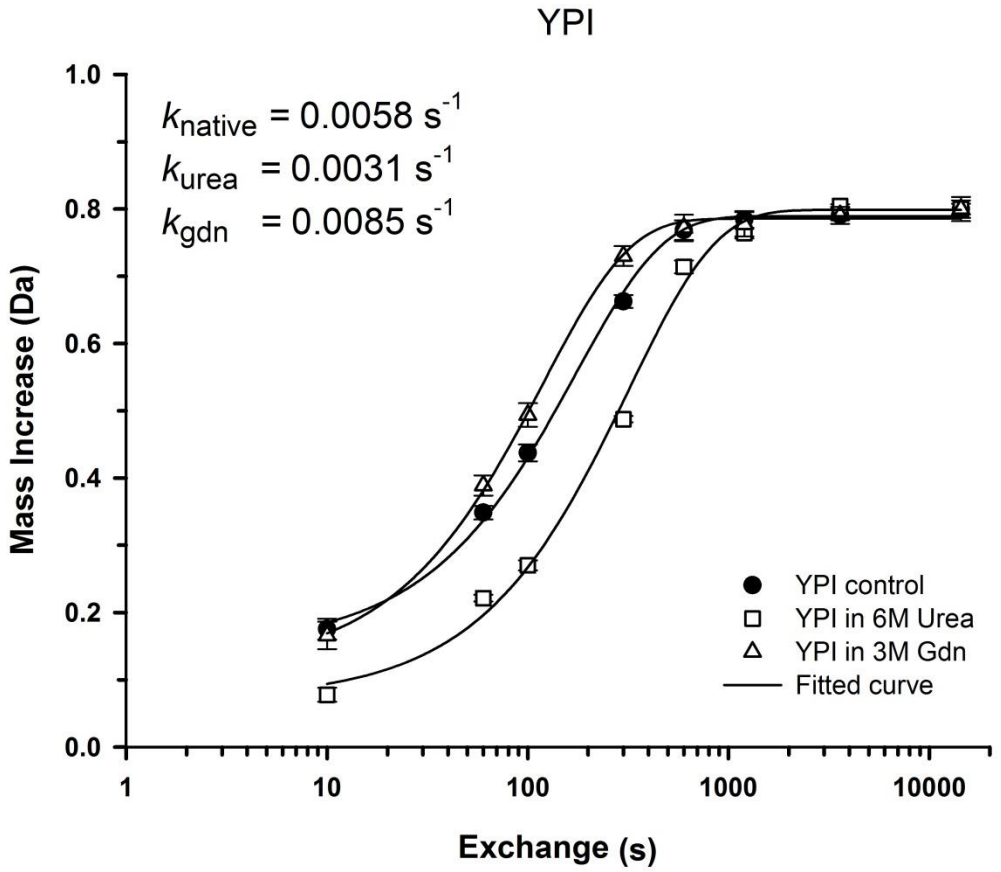


Figure 3.2. Hydrogen exchange for the unstructured model peptide (YPI) at three conditions: HX measured with no denaturant in citrate buffer at pD 6.5 (control) (filled points), in the presence of 6 M urea (open squares), and in the presence of 3 M guanidine (open triangles). The HX data were fit to equation 3 (see text for details) to obtain the rate constants for hydrogen exchange shown on the plot. The error bars denote the standard deviations from triplicate measurements.

Table 3.1. Fitting parameters and rate constants for YPI at three different conditions. HX measured with no denaturant in citrate buffer at pD 6.5 (control), in the presence of 6 M urea, and in the presence of 3 M guanidine. The fitting parameters were obtained using equation 3 (see text for details).

Condition	*y* (Da)	*a* (Da)	*k* (s^{-1})
No denaturants	0.15	0.64	0.0058
6 M urea	0.07	0.72	0.0031
3 M guanidine	0.11	0.67	0.0085

3.3.3 Peptic peptides as an unprotected reference state

Another way to prepare an unprotected reference state might be simply by labeling the peptides obtained by pepsin digestion of the proteins. We expect that the peptic reference peptides are unstructured because they are derived from already disordered proteins which have low hydrophobic and high polar and charged content of amino acid residues. This characteristic would make it difficult for the peptides to adopt secondary and tertiary structure. In this approach, the rate of chemical exchange is estimated based on an experimental determination of HX in an unprotected reference state. The hydrogen exchange labeling would occur using the same buffer system with the same composition which is an advantage point to this approach. The N-terminal segments of ACTR are completely unstructured under native conditions as revealed by NMR data.[(31)] Thus, if the HX by the peptic reference and the HX by the protein in the N-terminal region are the same, there would be good evidence that the peptide is an accurate unstructured reference. Unstructured internal standard peptides were added to confirm that the two states had consistent labeling. We carried out offline ACTR and CBP digestion to obtain the

peptides, as described in the Materials and Methods section. We carried out millisecond HX labeling on peptic peptides obtained from ACTR and CBP and compared the rates of HX with the rates of HX obtained from intact ACTR and CBP. Figure 3.3 shows representative deuterium uptake curves comparing HX by reference peptides with HX by two segments of ACTR and two segments of CBP measured under native conditions. The white filled squares denote the deuterium uptake by the proteins in unprotected reference state and the black filled diamonds denote the deuterium uptake by the proteins measured under native conditions. A complete set of deuterium uptake curves for ACTR and CBP in both states are provided in Supporting Information Figure 3.S3 and Figure 3.S4, respectively. To quantify the difference in HX between the reference and measured states for each segment of the proteins, we used the protection ratio, the ratio of $t_{50\%}$ of a peptide in the unprotected reference state to the $t_{50\%}$ of the measured state, as described in Materials and Methods section. An unstructured peptide should have a ratio of 1 because there is no protection meaning that the segment exchanges at the same rate as its reference peptide. If there is protection from exchange, then the protection ratio becomes greater than 1. A Table listing of all the peptides and their protection ratios is provided in Supporting Information (Table 3.S1). We have mapped all the protection ratios onto the primary sequence of ACTR and CBP as shown in Figure 3.4a and 3.5a.

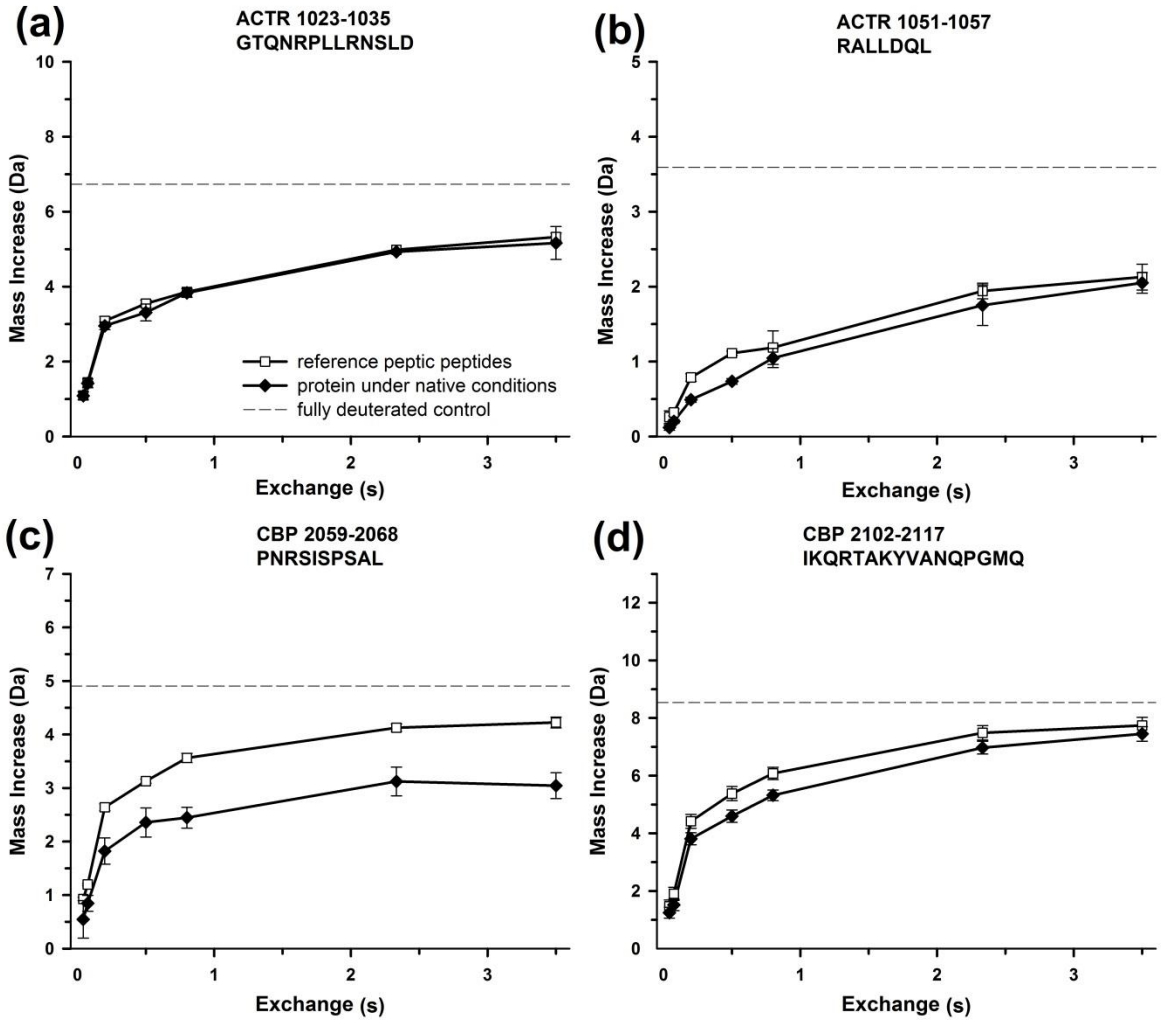


Figure 3.3. Deuterium uptake curves for the distinct regions from ACTR and CBP. (a) and (b) are representative peptides from ACTR and (c) and (d) are representative peptides from CBP. The horizontal axis in each plot is the exchange time in seconds and the vertical axis is the number of exchangeable amides by each peptide. The white filled squares represent uptake by the peptic reference peptides and the black filled diamonds represent uptake by the protein measured under native conditions. The error bars denote standard deviation from triplicate measurements. The horizontal dashed line represents the maximum deuterium uptake measured after exchange for 12 h.

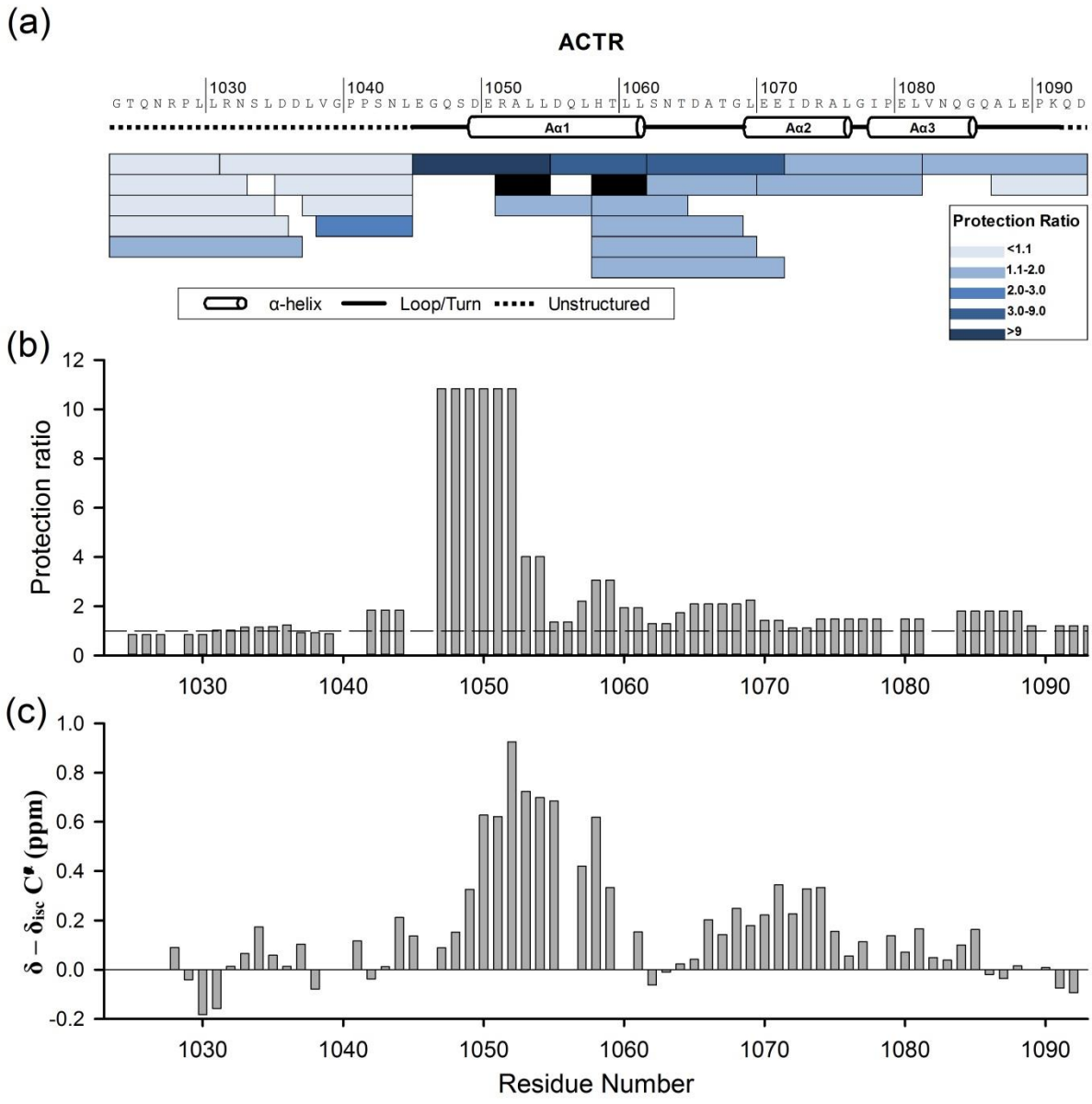


Figure 3.4. Measurement of residual helicity in ACTR. Segment-averaged protection ratio for each peptic peptide of ACTR was mapped onto the primary sequence and secondary structure (a) that ACTR forms in complex with its binding partner, CBP. [31] The protection ratios were calculated as described in the text. The protection ratios are indicated by the blue color scale with lighter blue represents lower protection. Two peptides in the middle region of ACTR, colored in black, were excluded from the calculations because they did not reach 50% exchange. Residue-averaged protection ratio obtained by averaging the peptide data (a) is shown in (b). The horizontal dashed line in (b) crossed the vertical axis at a value of 1 means there is no protection. Secondary chemical shift data obtained by NMR[31] is shown in (c).

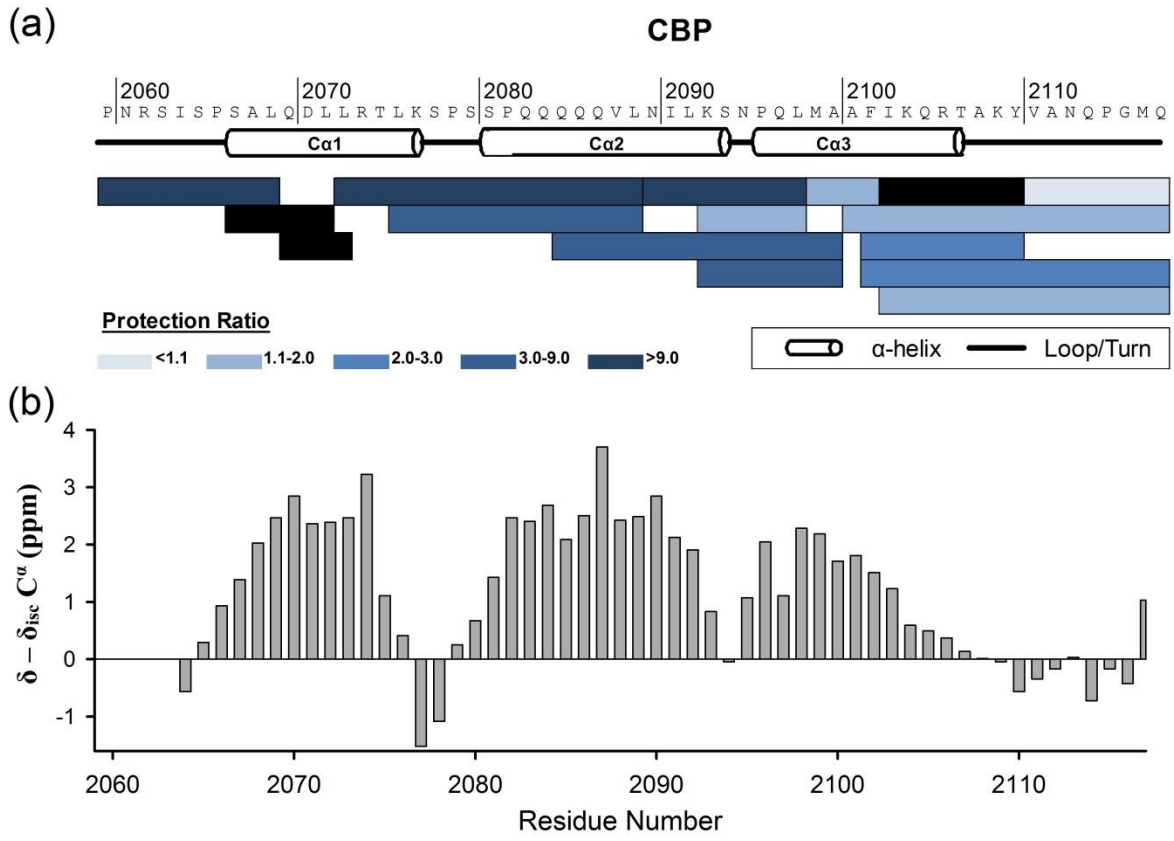


Figure 3.5. Measurement of residual helicity in CBP. Segment-averaged protection ratio for each peptic peptide of CBP was mapped onto the primary sequence and secondary structure (a). The protection ratios were calculated as described in the text. The protection ratios are indicated by the blue color scale with lighter blue represents lower protection. Three peptides of CBP, colored in black, were excluded from the calculations because they did not reach 50% exchange. Secondary chemical shift data obtained by NMR[39] is shown in (b). The plot was digitized using plot digitizer available online at http://arohatgi.info/WebPlotDigitizer/app/

For HX measured under native conditions, ACTR 1023-1035 segment exchange at the same rate as its peptide reference (see Figure 3.3a). This segment is located in the N-terminal region of ACTR which is an important region to validate the peptide reference state approach because NMR data show that this region is unstructured.[31] In contrast, we observed a significant protection from the exchange by ACTR 1051-1057 segment measured under native condition, as indicated by exchanging slower than its peptide reference (see Figure 3.3b). For the N-terminal

and C-terminal segments of ACTR, we observed no or slight difference with protection ratio less than 1.1. These segments cover 1023-1044 of the N-terminal region and 1087-1093 of the C-terminal of ACTR (see Figure 3.4a and Table 3.S1). The observed HX in these segments is consistent with the HX by their peptide references. However, in the middle regions of ACTR, several segments were significantly protected from HX. The segments cover 1038-1069 of the middle region of ACTR (see Figure 3.4a and Table 3.S1). Among these segments, there is only one segment with protection ratio as high as 11, which represents the highest degree of structure in ACTR. For ACTR, we could also calculate residue-averaged protection ratios because there were many overlapping peptides to calculate the protection for each residue. The residue-averaged data (see Figure 3.4b) is presented in the Discussion section.

For CBP, we observed substantial protection from exchange in CBP (2059-2068), located in the N-terminal region of the protein (see Figure 3.3c). This segment spans three residues of a region with the most intense value of NMR secondary chemical shift (see Figure 3.5b).[(39)] CBP 2102-2117 was also protected from the exchange as shown in Figure 3.3d, but the amount of protection is not as much in the N-terminal segment of CBP. Two segments (2066-2071 and 2069-2072) in the N-terminal region, colored in black (Figure 3.5a), are excluded from protection ratio calculation because they did not reach 50% exchange. For segment (2102-2109), there was not enough data in the HX by the segment measured under native conditions. In general, there is protection from exchange along all the length of CBP. The amount of protection varies from region to region. For example, the protection ratio for CBP 2072-2088, 2084-2099, and 2089-2097 segments were as high as 11 (see Figure 3.5 and Table 3.S1). In the C-terminal region of CBP, there is only one segment (CBP 2110-2117) with protection ratio equal to 1. With these data, we have been able to quantify the amount of protection in both ACTR and CBP.

3.4 Discussion

To shed some light on the mechanism of coupled folding and binding by IDPs and their interaction partners, we need to have a good understanding of the detailed structure of IDPs in the unbound state. It would be very difficult to investigate such a complex mechanism without knowing the IDPs conformations that may have transiently-folded conformers under native conditions. Previous work showed that millisecond HX can provide vital information about IDPs and their interactions.[23, 24, 26, 36, 45] It is essentially impossible to determine the degree of structure without having an accurate unprotected reference state. Our primary goal in this study was to develop a method that can accurately measure the degree of structure. For this purpose, we have tested several theoretical and experimental options to identify the best approach to generate an unprotected reference state for measuring the degree of structure in IDPs.

In the first attempt to have an accurate unprotected reference state, we used chemical exchange calculation as an unprotected reference state. We observed that ACTR and CBP peptides exchange significantly faster than the calculated exchange. Two model peptides, FKPGI and YPI, also exchanged faster than the calculated chemical exchange (see Figure 3.1). It is unlikely that this effect is due to structural protection of ACTR and CBP segments in their unstructured regions because we observe the same effect when we compared peptic reference peptides of the proteins and also for FKPGI peptide with their calculated chemical exchange (see Figure 3.1 and 3.S2). It is also unlikely that this effect is due to our quench-flow system because we also observe the differences in YPI exchange on seconds to hours timescale using more precise robotic labeling (see Figure 3.1d). Most of the ACTR and CBP segments exchange more quickly than the calculated exchange have residues such as alanine, leucine, and arginine with the least helical propensity.[42] The two model peptides also contain residues with very poor

helical propensity like proline, isoleucine, and lysine[42] and are short peptides. Previously, Del Mar et al. reported that many segments across regions of α-synuclein, an intrinsically disordered protein, exchange at rates faster than the calculated chemical exchange rate.[28] Based on what we found for the calculated exchange and what others previously observed[28] the chemical exchange calculation is not accurate enough to measure protection in IDPs. The chemical exchange calculations are based on measured HX rates obtained using model dipeptides, oligopeptides, and polypeptides at the high ionic strength (0.5 M KCl).[27] The calculation accounts for the inductive and steric effects of the neighboring side chains but longer range effects might also be present. So, in many cases, this calculation may not be reliable for different solution conditions, for example, low ionic strength or high concentrations of additives such as guanidine and urea in buffers. These factors may lead to larger deviations from ideal solution behavior. Some other factors need to be accounted for in the HX. These include the labeling ratio as we have 1:5 sample-to-labeling buffers in this experiment and also other electrostatic field effects of side chains of disordered proteins. All these factors may have led to the acceleration of the chemical exchange for the ACTR, CBP, and the model peptides. Thus, if the protection ratio were determined using chemical exchange calculation, it would underestimate the degree of protection. The data would become misleading about the structure of IDPs. In contrast, the error in the degree of protection for well-folded proteins might become unimportant because the difference between the measured rate of exchange and the rate of chemical exchange is often many orders of magnitude. Therefore, a better choice is to find a way to measure chemical exchange experimentally and directly for IDPs.

We tested an experimental method to completely unfold the proteins using urea and guanidine to produce an unprotected reference state. However, we observed a significant

decrease in HX rate of the model peptide (YPI) in the presence of urea and an opposite effect in the presence of guanidine (see Figure 3.2). We found that the rate of exchange in 6 M urea was 1.87 slower than the control condition and the exchange percentage decreased from 48% down to 28% in the presence of urea when compared at $t_{50\%}$ for control YPI. In contrast, the rate of exchange in 3 M guanidine was 1.47 faster than the control condition and the exchange percentage increased from 48% up to 63% in the presence of guanidine. Urea can form hydrogen bonds with NH group of peptides thus leading to protection from HX due to changes in the solution conditions that alter the rate of chemical exchange.[46] Guanidine has some impact on the acid-catalyzed peptide hydrogen exchange.[46] A theoretical correction of the data might be able to compensate for the differences raised by the difference in buffer compositions. For example, Zhang et al. used an internal standard peptide (PPPI) in different guanidine concentrations for this purpose.[29] However, the data correction may not be reliable enough to map degree of structure across all the regions of IDPs because they are only weakly protected at most.

We found that peptic reference peptide for millisecond HX is the best method to quantify the degree of protection. Before validating this approach, we employed two model peptides as internal standards to confirm that the experiments have consistent labeling. The results show that the rates of HX are the same in both states confirming good labeling control (see Figure 3.S5). We observed similar rates of HX with protection ratio close to 1 for the N-terminal segments of ACTR (1023-1034, and 1023-1036) which showed no residual helicity by NMR secondary chemical shift[31] (see Figures 3.4c and 3.S3). Our data demonstrate that the peptic reference peptide method can work as a good unprotected reference state. Because any difference in the composition of the buffer may result in the differences in the exchange rate, one of the major

advantages of using peptic reference peptide as an unprotected reference state is that there is no need to change the composition of the buffers. The other advantage is that no synthetic peptide standard is needed.

There are extensive NMR data characterizing residual structure in ACTR in both the isolated form and in complex with CBP. [11, 31, 33, 39, 47] On the basis of the NMR data, ACTR has the characteristics of an unstructured random coil that transiently folds in the regions that form helices in the complex with CBP. Because ACTR has both unstructured and transiently structured regions, it is a good model to determine the validity of an unprotected reference state. In general, we found good agreement between the millisecond HX and NMR secondary chemical shifts for measuring residual structure in ACTR. The highest percentage of residual helicity is located in helix 1 when the NMR data of ACTR is measured relative to reference shifts measured in 6 M urea (see Figure 3.4c).[31] Although urea might have effects on chemical HX, NMR chemical shifts only depend on the protein structure. Our millisecond HX data are in good agreement with NMR data.

In the N-terminal region of ACTR, we observed the lowest segment-averaged protection ratios. To calculate protection ratio at near single-residue resolution, a weighted residue-by-residue average of the segment-averaged protection ratio was used. We mapped the residue-averaged protection ratio on the amino acid sequence of ACTR. We observed a very good agreement for the N-terminal region of ACTR between the residue-averaged protection ratio and NMR data (see Figures 3.4b and 3.4c). The same agreement was observed for the C-terminal region of ACTR. Although the residue-averaged protection ratio showed the highest residual structure in helix 1 region and is in good agreement with NMR data (see Figures 3.4b and 3.4c), the t50% value is displaced more toward the N-terminal region. The reason for this displacement

is that the number of segments accounted for with the residue-averaged protection ratio calculation is low. There is only one segment that spans residues 1045-1051 leading to decreased spatial resolution in this region. In addition, two other segments, 1051-1054 and 1058-1061, were excluded from residue-averaged protection ratio calculation because they did not reach 50% exchange (see Table 3.S1). These two peptides (RALL, HTLL) are short segments that end with two leucine residues that slow down chemical hydrogen exchange of the amide by the side chain steric effect.[27]

The apparent structure across most regions of CBP, as shown in our data (see Figure 3.5a), is consistent with the NMR data. We only used segment-averaged protection ratio because the number of peptides for CBP was limited which makes the residue averaging calculation unreliable. Unlike ACTR, CBP has the characteristics of a molten globule that transiently becomes unstructured.[23, 39-41] In general, we find a good correlation between the segment averaged protection ratio with NMR secondary chemical shift data.[39] The C-terminal region of CBP was the least protected, with protection ratios ranging from 1 to 3 (see Figure 3.5a, and Table 3.S1). These data are in a good agreement with NMR data in the helix 3 region. However, we see some discrepancy in the N-terminal region because there are not enough segments to cover only unstructured the residues of CBP, 2059-2067 and 2079-2085. There are two other segments, 2066-2071 and 2069-2072, in this region that are excluded from the protection ratio calculation because they did not reach 50% exchange on this time scale. These segments, SALQDL and QDLL, also have leucine residues at the end of the segment which slow down the chemical exchange.

Our approach for mapping the degree of IDPs structure relies on the use of peptic reference peptides as an unprotected reference state in the HX experiment. This approach can

offer a simple, easy to apply, and reliable empirical unprotected reference state. There is clearly no need to change buffer composition by adding a chemical denaturant to obtain an unprotected reference state. Data normalization due to altered buffer composition is not required making it less susceptible to HX miscalculations. We successfully mapped the degree of protection in two different model IDPs, ACTR and CBP. This approach can give a new insight into the analysis of transient structure in other IDPs. However, this approach may not work for all kinds of proteins because if a peptide can form structure even after protein digestion then it cannot act as an unprotected reference peptide. This could be a problem for globular proteins that have a much stronger tendency to fold. In such cases, peptic peptides in a low denaturant concentration might be a better choice.

3.5 References

1. Sigler, P. B. (1988) Transcriptional activation. Acid blobs and negative noodles, *Nature 333*, 210-212.

2. Garner, E., Cannon, P., Romero, P., Obradovic, Z., and Dunker, A. K. (1998) Predicting Disordered Regions from Amino Acid Sequence: Common Themes Despite Differing Structural Characterization, *Genome informatics. Workshop on Genome Informatics 9*, 201-213.

3. Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions?, *Proteins 41*, 415-427.

4. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) Intrinsic disorder and protein function, *Biochemistry 41*, 6573-6582.

5. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nature reviews. Molecular cell biology 6*, 197-208.

6. Tompa, P., Fuxreiter, M., Oldfield, C. J., Simon, I., Dunker, A. K., and Uversky, V. N. (2009) Close encounters of the third kind: disordered domains and the interactions of proteins, *BioEssays : news and reviews in molecular, cellular and developmental biology 31*, 328-335.

7. Tantos, A., Han, K.-H., and Tompa, P. (2012) Intrinsic disorder in cell signaling and gene transcription, *Molecular and cellular endocrinology 348*, 457-465.

8. Radivojac, P., Iakoucheva, L. M., Oldfield, C. J., Obradovic, Z., Uversky, V. N., and Dunker, A. K. (2007) Intrinsic disorder and functional proteomics, *Biophysical journal 92*, 1439-1456.

9. Kovacs, D., Szabo, B., Pancsa, R., and Tompa, P. (2013) Intrinsically disordered proteins undergo and assist folding transitions in the proteome, *Archives of biochemistry and biophysics 531*, 80-89.

10. Wright, P. E., and Dyson, H. J. (2009) Linking folding and binding, *Current opinion in structural biology 19*, 31-38.

11. Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J., Evans, R. M., and Wright, P. E. (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators, *Nature 415*, 549-553.

12. Lacy, E. R., Filippov, I., Lewis, W. S., Otieno, S., Xiao, L., Weiss, S., Hengst, L., and Kriwacki, R. W. (2004) p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding, *Nature structural & molecular biology 11*, 358-364.

13. Uversky, V. N. (2011) Intrinsically disordered proteins from A to Z, *The International journal of biochemistry & cell biology 43*, 1090-1103.

14. Receveur-Brechot, V., Bourhis, J. M., Uversky, V. N., Canard, B., and Longhi, S. (2006) Assessing protein disorder and induced folding, *Proteins 62*, 24-45.

15. Eliezer, D. (2009) Biophysical characterization of intrinsically disordered proteins, *Current opinion in structural biology 19*, 23-30.

16. Scott, D. J., and Winzor, D. J. (2015) Characterization of Intrinsically Disordered Proteins by Analytical Ultracentrifugation, *Methods in enzymology 562*, 225-239.

17. Chandrashekaran, I. R., Mohanty, B., Linossi, E. M., Dagley, L. F., Leung, E. W., Murphy, J. M., Babon, J. J., Nicholson, S. E., and Norton, R. S. (2015) Structure and Functional Characterization of the Conserved JAK Interaction Region in the Intrinsically Disordered N-Terminus of SOCS5, *Biochemistry 54*, 4672-4682.

18. Hellman, M., Tossavainen, H., Rappu, P., Heino, J., and Permi, P. (2011) Characterization of intrinsically disordered prostate associated gene (PAGE5) at single residue resolution by NMR spectroscopy, *PloS one 6*, e26633.

19. Smaldone, G., Falanga, A., Capasso, D., Guarnieri, D., Correale, S., Galdiero, M., Netti, P. A., Zollo, M., Galdiero, S., Di Gaetano, S., and Pedone, E. (2013) gH625 is a viral derived peptide for effective delivery of intrinsically disordered proteins, *International journal of nanomedicine 8*, 2555-2565.

20. Marcsisin, S. R., and Engen, J. R. (2010) Hydrogen exchange mass spectrometry: what is it and what can it tell us?, *Analytical and bioanalytical chemistry 397*, 967-972.

21. Zhu, S., Shala, A., Bezginov, A., Sljoka, A., Audette, G., and Wilson, D. J. (2015) Hyperphosphorylation of Intrinsically Disordered Tau Protein Induces an Amyloidogenic Shift in Its Conformational Ensemble, *PloS one 10*, e0120416.

22. Englander, S. W., and Kallenbach, N. R. (1983) Hydrogen exchange and structural dynamics of proteins and nucleic acids, *Quarterly reviews of biophysics 16*, 521-655.

23. Keppel, T. R., Howard, B. A., and Weis, D. D. (2011) Mapping Unstructured Regions and Synergistic Folding in Intrinsically Disordered Proteins with Amide H/D Exchange Mass Spectrometry, *Biochemistry 50*, 8722-8732.

24. Keppel, T. R., and Weis, D. D. (2013) Analysis of disordered proteins using a simple apparatus for millisecond quench-flow H/D exchange, *Analytical chemistry 85*, 5161-5168.

25. Goswami, D., Devarakonda, S., Chalmers, M. J., Pascal, B. D., Spiegelman, B. M., and Griffin, P. R. (2013) Time window expansion for HDX analysis of an intrinsically disordered protein, *Journal of the American society for mass spectrometry 24*, 1584-1592.

26. Rob, T., Liuni, P., Gill, P. K., Zhu, S., Balachandran, N., Berti, P. J., and Wilson, D. J. (2012) Measuring dynamics in weakly structured regions of proteins using microfluidics-enabled subsecond H/D exchange mass spectrometry, *Analytical Chemistry 84*, 3771-3779.

27. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange, *Proteins: structure, function, and bioinformatics 17*, 75-86.

28. Del Mar, C., Greenbaum, E. A., Mayne, L., Englander, S. W., and Woods, V. L., Jr. (2005) Structure and properties of alpha-synuclein and other amyloids determined at the amino acid level, *Proceedings of the national academy of sciences of the United States of America 102*, 15477-15482.

29. Zhang, Z., Zhang, A., and Xiao, G. (2012) Improved protein hydrogen/deuterium exchange mass spectrometry platform with fully automated data processing, *Analytical chemistry 84*, 4942-4949.

30. Keppel, T. R., Jacques, M. E., Young, R. W., Ratzlaff, K. L., and Weis, D. D. (2011) An efficient and inexpensive refrigerated LC system for H/D exchange mass spectrometry, *Journal of the American society for mass spectrometry 22*, 1472-1476.

31. Kjaergaard, M., Norholm, A. B., Hendus-Altenburger, R., Pedersen, S. F., Poulsen, F. M., and Kragelund, B. B. (2010) Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II?, *Protein science : a publication of the protein society 19*, 1555-1564.

32. Kjaergaard, M., Iešmantavičius, V., and Poulsen, F. M. (2011) The interplay between transient α-helix formation and side chain rotamer distributions in disordered proteins probed by methyl chemical shifts, *Protein science : a publication of the protein society 20*, 2023-2034.

33. Iesmantavicius, V., Jensen, M. R., Ozenne, V., Blackledge, M., Poulsen, F. M., and Kjaergaard, M. (2013) Modulation of the intrinsic helix propensity of an intrinsically disordered protein reveals long-range helix-helix interactions, *Journal of the American chemical society 135*, 10155-10163.

34. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) Measurement of protein using bicinchoninic acid, *Analytical biochemistry 150*, 76-85.

35. Connelly, G. P., Bai, Y., Jeng, M. F., and Englander, S. W. (1993) Isotope effects in peptide group hydrogen exchange, *Proteins 17*, 87-92.

36. Keppel, T. R., and Weis, D. D. (2015) Mapping residual structure in intrinsically disordered proteins at residue resolution using millisecond hydrogen/deuterium exchange and residue averaging, *Journal of the American society for mass spectrometry 26*, 547-554.

37. Wang, L., Pan, H., and Smith, D. L. (2002) Hydrogen exchange-mass spectrometry: optimization of digestion conditions, *Molecular & cellular proteomics : MCP 1*, 132-138.

38. Busby, S. A., Chalmers, M. J., and Griffin, P. R. (2007) Improving digestion efficiency under H/D exchange conditions with activated pepsinogen coupled columns, *International journal of mass spectrometry 259*, 130-139.

39. Ebert, M. O., Bae, S. H., Dyson, H. J., and Wright, P. E. (2008) NMR relaxation study of the complex formed between CBP and the activation domain of the nuclear hormone receptor coactivator ACTR, *Biochemistry 47*, 1299-1308.

40. Kjaergaard, M., Teilum, K., and Poulsen, F. M. (2010) Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP, *Proceedings of the national academy of sciences of the United States of America 107*, 12535-12540.

41. Lin, C. H., Hare, B. J., Wagner, G., Harrison, S. C., Maniatis, T., and Fraenkel, E. (2001) A Small Domain of CBP/p300 Binds Diverse Proteins: Solution Structure and Functional Studies, *Molecular cell 8*, 581-590.

42. Nick Pace, C., and Martin Scholtz, J. (1998) A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins, *Biophysical journal 75*, 422-427.

43. Hvidt, A., and Nielsen, S. O. (1966) Hydrogen exchange in proteins, *Advances in protein chemistry 21*, 287-386.

44. Bennion, B. J., and Dagget, V. (2003) The molecular basis for the chemical denaturation of proteins by urea.(Author Abstract), *Proceedings of the national academy of sciences of the United States 100*, 5142.

45. Resetca, D., and Wilson, D. J. (2013) Characterizing rapid, activity-linked conformational transitions in proteins via sub-second hydrogen deuterium exchange mass spectrometry, *The FEBS journal 280*, 5616-5625.

46. Lim, W. K., Rosgen, J., and Englander, S. W. (2009) Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group, *Proceedings of the national academy of sciences of the United States of America 106*, 2595-2600.

47. Kjaergaard, M., Brander, S., and Poulsen, F. M. (2011) Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH, *Journal of Biomolecular NMR 49*, 139-149.

Table 3.S1. ACTR and CBP peptic peptides $t_{50\%}$ values at each state and their protection ratio.

protein	residues	sequence	$t_{50\%}$ (s)		$t_{50\%}$ ratio
			reference	native	
ACTR	1023-1030	GTQNRPLL	0.31	0.19	0.63
ACTR	1023-1032	GTQNRPLLRN	0.3	0.24	0.8
ACTR	1023-1034	GTQNRPLLRNSL	0.19	0.21	1.09
ACTR	1023-1035	GTQNRPLLRNSLD	0.39	0.41	1.06
ACTR	1023-1036	GTQNRPLLRNSLDD	0.42	0.6	1.45
ACTR	1031-1044	RNSLDDLVGPPSNL	1.32	1.4	1.06
ACTR	1035-1044	DDLVGPPSNL	1.46	1.29	0.88
ACTR	1037-1044	LVGPPSNL	0.32	0.27	0.85
ACTR	1038-1044	VGPPSNL	0.17	0.45	2.7
ACTR	1045-1054	EGQSDERALL	0.15	1.6	10.84
ACTR	1051-1057	RALLDQL	1.96	2.66	1.36
ACTR	1055-1061	DQLHTLL	0.21	0.65	3.05
ACTR	1058-1064	HTLLSNT	0.36	0.4	1.11
ACTR	1058-1068	HTLLSNTDATG	0.33	0.42	1.29
ACTR	1058-1069	HTLLSNTDATGL	0.47	0.91	1.93
ACTR	1058-1071	HTLLSNTDATGLEE	0.71	1.02	1.43
ACTR	1062-1069	SNTDATGL	0.3	0.9	3.04
ACTR	1062-1071	SNTDATGLEE	0.74	1.06	1.43
ACTR	1070-1081	EEIDRALGIPEL	3.31	3.7	1.12
ACTR	1072-1081	IDRALGIPEL	1.84	3.11	1.7
ACTR	1082-1093	VNQGQALEPKQD	0.57	1.03	1.8
ACTR	1087-1093	ALEPKQD	1.82	1.97	1.08
CBP	2059-2068	PNRSISPSAL	0.18	1.86	10.08
CBP	2072-2088	LRTLKSPSSPQQQQQVL	0.18	1.78	9.67
CBP	2075-2088	LKSPSSPQQQQQVL	0.19	0.96	4.94
CBP	2084-2099	QQQVLNILKSNPQLMA	0.19	1.61	8.51
CBP	2089-2097	NILKSNPQL	0.28	3.14	11.32
CBP	2092-2097	KSNPQL	1.08	1.83	1.69
CBP	2092-2099	KSNPQLMA	0.51	1.82	3.57
CBP	2098-2101	MAAF	1.63	2.58	1.58
CBP	2100-2117	AFIKQRTAKYVANQPGMQ	0.38	0.59	1.53
CBP	2101-2109	FIKQRTAKY	0.17	0.34	2.06
CBP	2101-2117	FIKQRTAKYVANQPGMQ	0.19	0.47	2.42
CBP	2102-2117	IKQRTAKYVANQPGMQ	0.19	0.37	1.9
CBP	2110-2117	VANQPGMQ	0.2	0.2	0.98

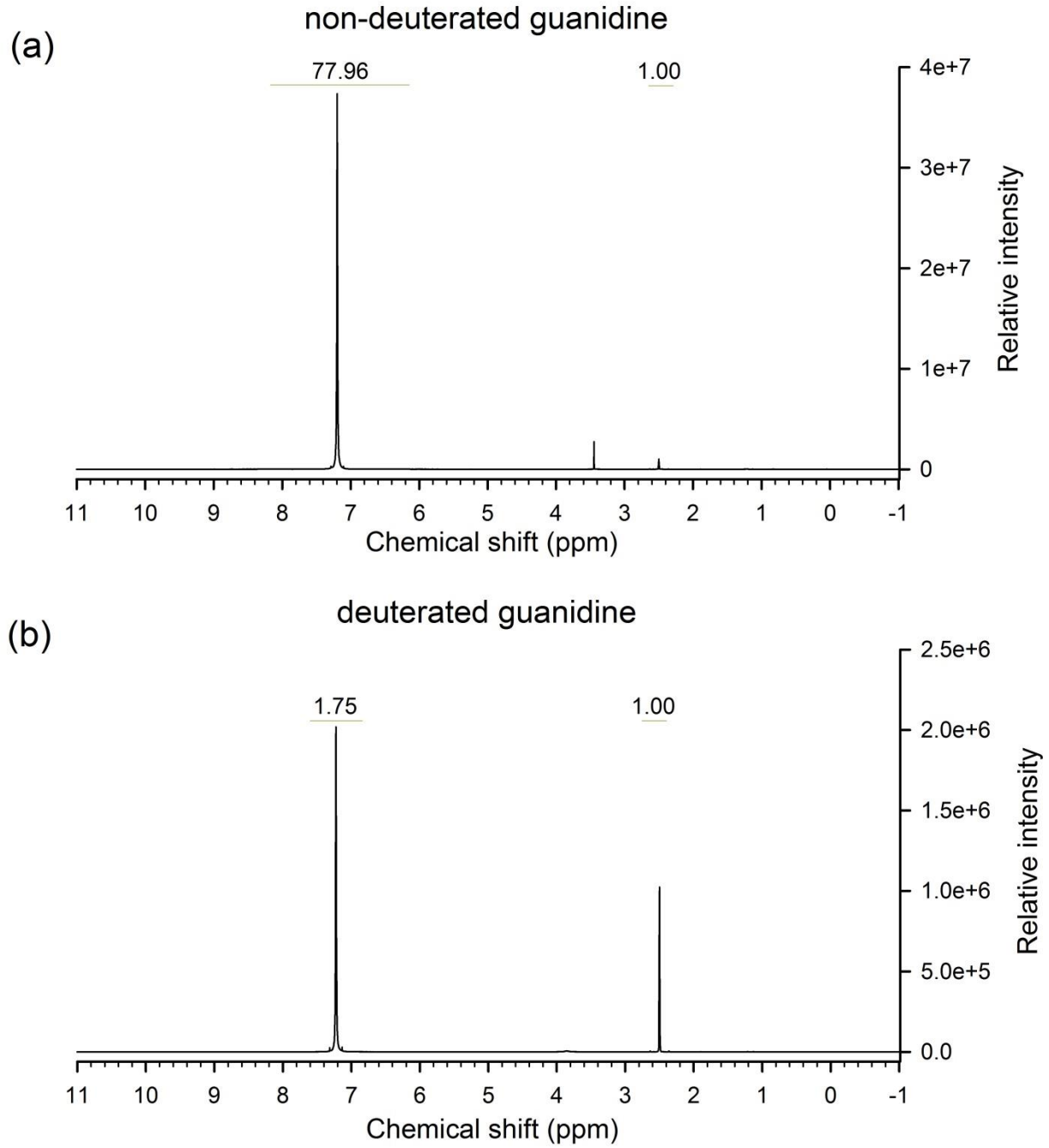


Figure 3.S1. Deuteration of guanidine. ^{1}H NMR data for (a) guanidine hydrochloride (control) and (b) guanidine after complete deuteration in D_2O (99.9 %) at room temperature with two rounds of dissolution/lyophilization as described in the text. The deuteration percentage was calculated after normalization the integration value of the solvent (DMSO) proton peak (2.5 ppm) in both conditions to 1.0. The guanidine deuteration percentage (97 %D) was then calculated using the number of protons (peak 7.2 ppm) in deuterated guanidine and the number of protons in non-deuterated control.

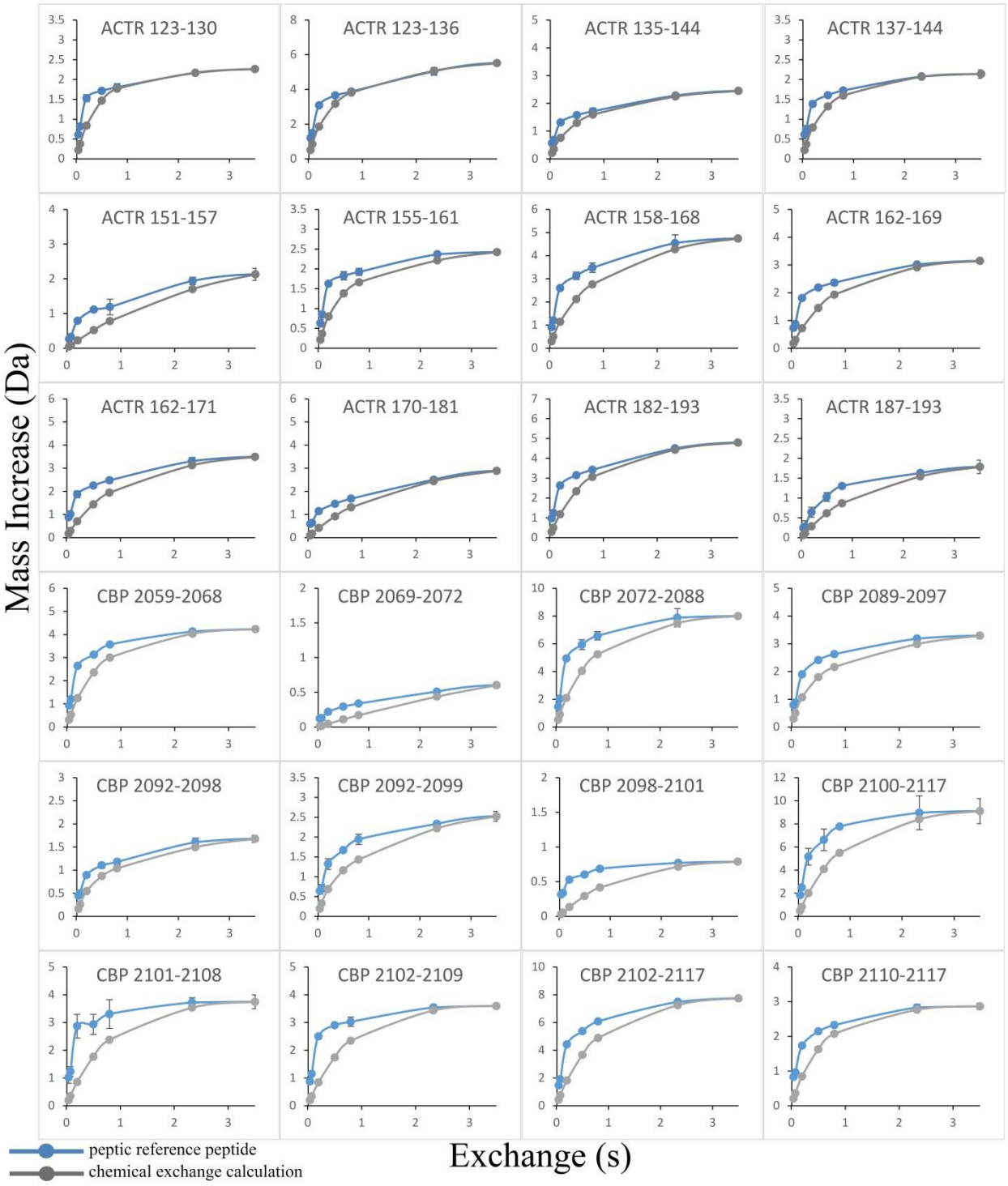


Figure 3.S2. Hydrogen exchange by peptides from ACTR and CBP. The blue lines represent uptake by the peptic reference peptides and the gray lines represent the chemical exchange calculation normalized to match the latest measured time point of each respective curve as described in the text.

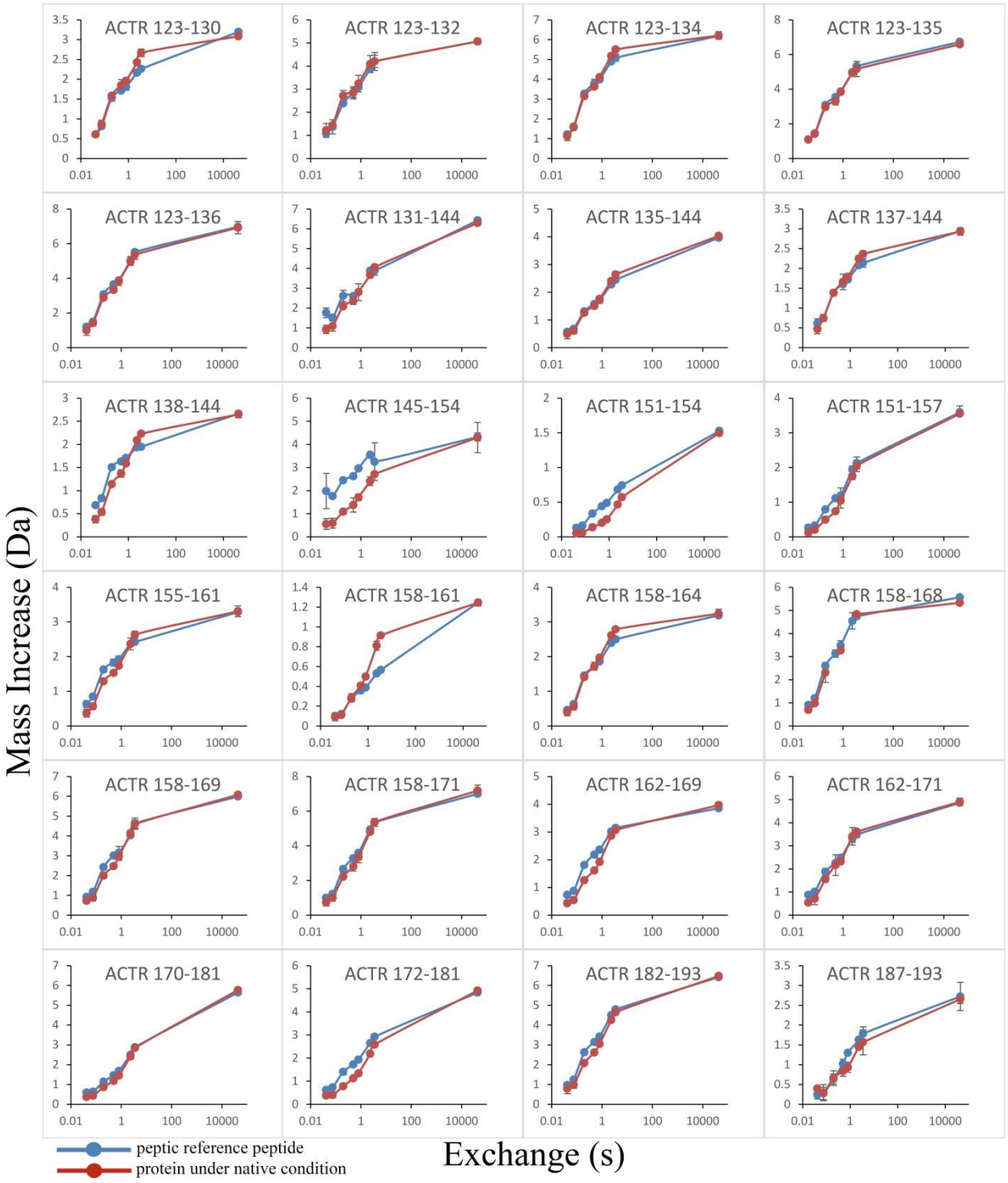


Figure 3.S3. Hydrogen exchange by peptides from ACTR. The blue lines represent uptake by the peptic reference peptides and the red lines represent uptake by the protein measured under native conditions.

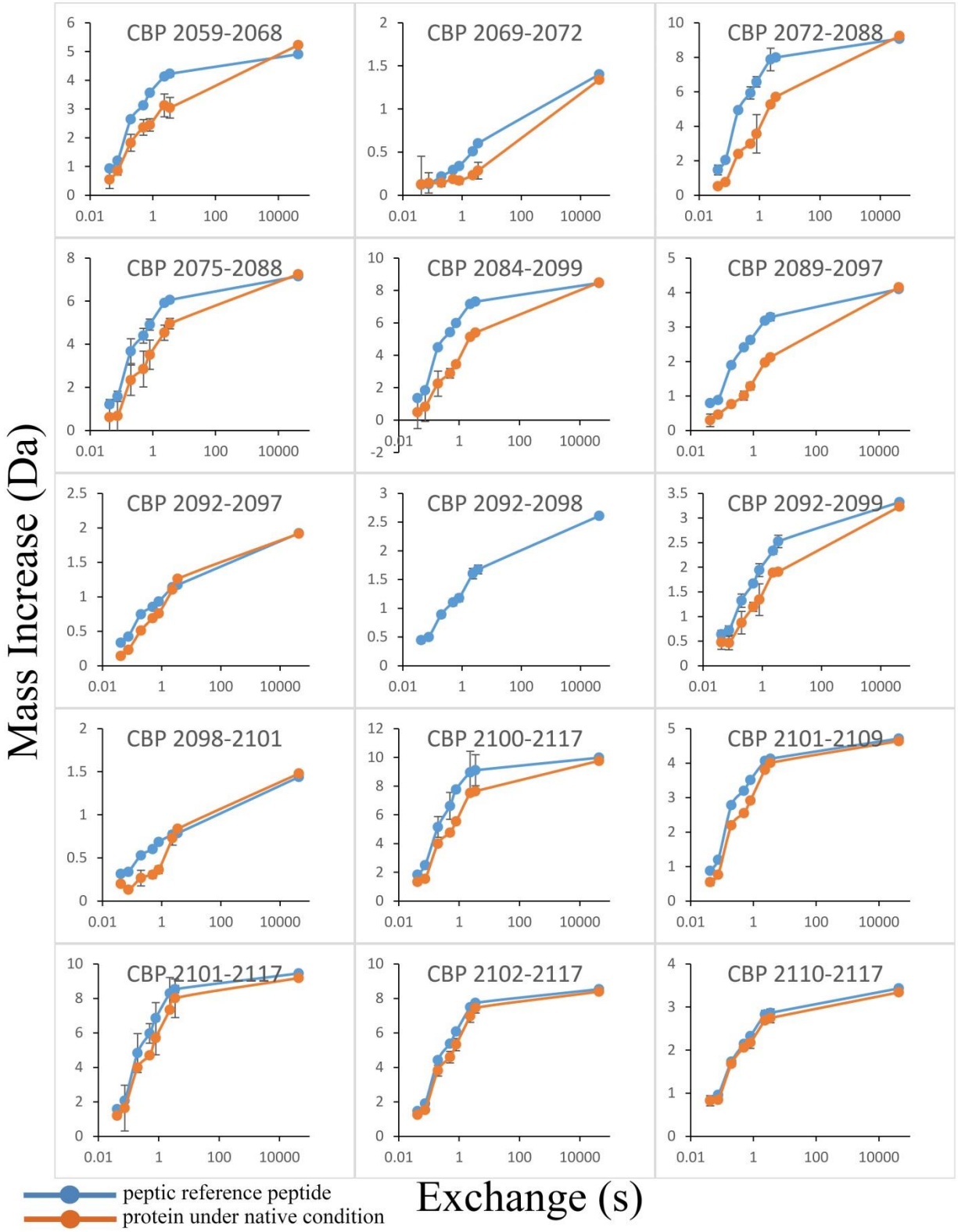


Figure 3.S4. Hydrogen exchange by peptides from CBP. The blue lines represent uptake by the peptic reference peptides and the orange lines represent uptake by the protein measured under native conditions.

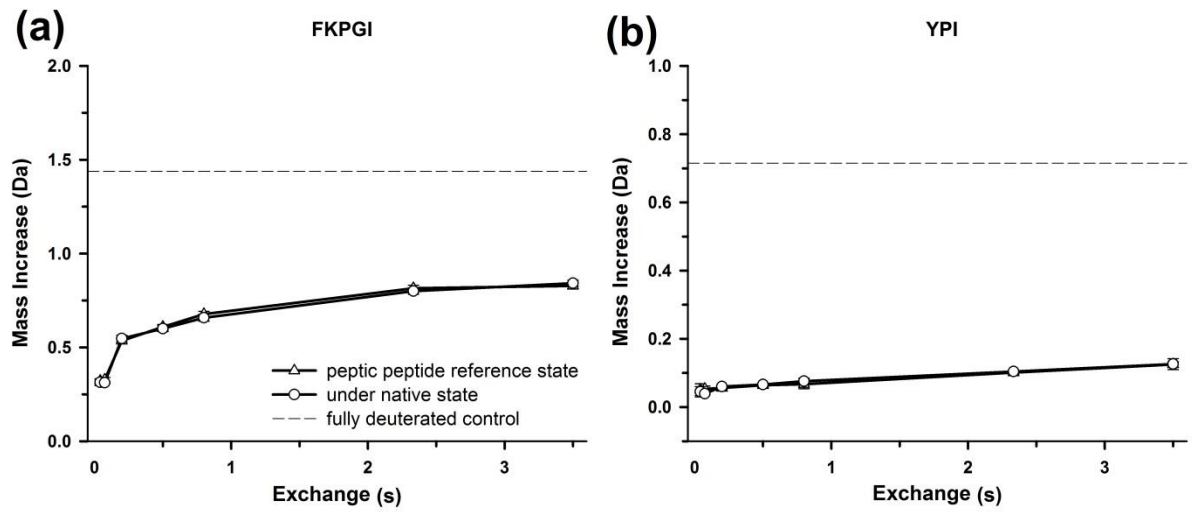


Figure 3.S5. Deuterium uptake curves for two unstructured model peptides, (a) FKPGI and (b) YPI. The open triangles represent uptake by the peptide in reference states and the open circles represent uptake by the peptides measured under native conditions. The error bars denote standard deviation from triplicate measurements. The horizontal axis in each plot is the exchange time in seconds and the vertical axis is the number of exchangeable amides by each peptide. The horizontal dashed line represents the maximum deuterium uptake measured after exchange for 12 h.

Chapter four

Application of fast photochemical oxidation of proteins (FPOP) to study intrinsically disordered proteins and their folded complexes

4.1 Introduction

Intrinsically disordered proteins (IDPs) have regions that are highly flexible and lack stable secondary or tertiary structure.[1-3] Although many studies have shown that IDPs participate in functional interactions, less is known about the structural details of the interactions.[4, 5] Since IDPs are unfolded or partially unfolded and are very flexible, they have a high solvent accessible surface. These properties of IDPs represent a challenge for some biophysical techniques to analyze this class of proteins.

Over the past two decades, mass spectrometry (MS) has become a powerful technique for biophysical characterization of proteins due to its high sensitivity and variety of choices for sample preparation and instrumentation.[6-8] One of the mass spectrometry based techniques used for IDPs conformation and dynamic studies is hydrogen exchange mass spectrometry (HX-MS).[9-11] See Section 1.3.3 for details about HX-MS. Although HX-MS has shown promise in analysis of folded proteins and their complexes, the application of conventional HX is still limited in IDPs studies. The reason is because of highly flexible and rapidly fluctuating secondary and tertiary structure properties of IDPs which results in complete exchange prior to the first HX measurements. More sophisticated HX techniques are required in order to probe the structural change in IDPs. One of these techniques is by using millisecond HX[10, 12] (see Section 1.3.3.3 and Chapter 2 of this Dissertation). In general, HX-MS can provide an average picture of the protein conformational changes that are induced by mutation, aggregation, and ligand and partner protein binding.[13-15] On the other hand, HX has some limitations. One of the limitations is the requirement of specialized LC-MS-based strategies to preserve the deuterium label of a protein.[16] In HX experiments, it is required to minimize back-exchange (deuterium label for protons) because higher deuterium loss causes loss of valuable information about protein

conformation and dynamics.[16, 17] In addition, HX-MS provides structural information based on amide backbone hydrogens but does not probe side chain residues. Finally, the spatial resolution in a typical HX-MS experiment is highly dependent on the proteolytic digestion step prior to MS analysis.[16] HX-MS can provide localized structural information at a resolution of 5-10 amino acid residues. Therefore, there is a need for other techniques that lack the limitations mentioned above and have better resolution for analyzing IDPs.

In this work, we evaluate the application of another MS-based protein footprinting technique to study the conformational change of an IDP upon binding to a partner IDP. The technique is fast photochemical oxidation of proteins (FPOP). The FPOP technique is based on covalent labeling of amino acid side chains of a protein with hydroxyl radicals generated by H_2O_2 photolysis into two hydroxyl radicals by an excimer laser (λ=248) that leads to oxidative modification of the surface accessible residues[18, 19] (see Section 1.3.4.2 for more details). This technique is based on probing solvent accessibility of amino acid side chains by modifying the side chains, and such modification is irreversible and can be preserved throughout the subsequent steps: proteolysis and LC-MS/MS analysis.[16, 18, 20] Single amino acid resolution can be achieved with the use of this technique. Although the hydroxyl radical reaction with side chains is non-specific, the reactivity of labeling side chains is not equal. Table 4.1 lists rate constants of the reaction of different amino acid side chains with hydroxyl radicals.[21] As shown in the table, hydroxyl radicals label side chains with different efficiencies. To explore the applicability of this technique for studying IDPs, we used ACTR and CBP. ACTR and CBP are useful models for conformational analysis of IDPs. On their own, CBP is a mostly helical molten globule while ACTR is unstructured.[22-24] The proteins bind with high affinity to form a well-defined complex with three helices each.[22, 23, 25] Our previous work showed that HX can provide important

information about IDPs, but only when labeling takes place on the millisecond-second timescale.[12] Since protein folding is slower than microseconds, fast photochemical oxidation of proteins (FPOP) might reveal states of IDPs that are undetectable by HX. Here, we used FPOP technique to study ACTR and CBP in their free and bound forms. The data show that FPOP can provide valuable information to compare two states of IDPs. The data illustrate the usefulness and limitations of FPOP-MS analysis to characterize and localize regions of protein-protein interactions involving intrinsically disordered proteins.

Table 4.1. Rate constants of the reactions of hydroxyl radical with amino acid side chains at corresponding pHs.[21, 28]

amino acid	hydroxyl radical	
	rate ($M^{-1} s^{-1}$)	pH
Cys (C)	3.5×10^{10}	7
Trp (W)	1.3×10^{10}	6.5-8.5
Tyr (Y)	1.3×10^{10}	7
Met (M)	8.5×10^{9}	6-7
Phe (F)	6.9×10^{9}	7-8
His (H)	4.8×10^{9}	7.5
Arg (R)	3.5×10^{9}	6.5-7.5
Cystine	2.1×10^{9}	6.5
Ile (I)	1.8×10^{9}	6.6
Leu (L)	1.7×10^{9}	~ 6
Val (V)	8.5×10^{8}	6.9
Pro (P)	6.5×10^{8}	6.8
Gln (Q)	5.4×10^{8}	6.0
Thr (T)	5.1×10^{8}	6.6
Lys (K)	3.5×10^{8}	6.6
Ser (S)	3.2×10^{8}	~ 6
Glu (E)	2.3×10^{8}	6.5
Ala (A)	7.7×10^{7}	5.8
Asp (D)	7.5×10^{7}	6.9
Asn (N)	4.9×10^{7}	6.6
Gly (G)	1.7×10^{7}	5.9

4.2 Materials and methods

All of the FPOP labeling and mass spectrometry analyses for this project were performed in the NIH/NIGMS Biomedical Mass Spectrometry Resource at the School of Medicine at Washington University in St. Louis.

4.2.1 Materials

ACTR$^{1023\text{-}1093}$ (UniProt NCOA3_HUMAN) and CBP$^{2059\text{-}2117}$ (UniProt CBP_MOUSE) domains were co-expressed from a pET22B co-expression vector[(22)] in *Escherichia coli* BL21 (DE3) and separately purified[(9)] (described in Section 3.2.1). The concentrations of ACTR and CBP were 60 µM and 17.2 µM, respectively. Formic acid, hydrogen peroxide (30 %), trypsin, *L*-glutamine, *L*-methionine, catalase, sodium phosphate, sodium chloride, tris-hydrochloride, and HPLC-grade solvents were all purchased from Sigma Aldrich (St. Louis, MO). Nu-Tip C-18 zip tips were obtained from Glygen Corporation (Columbia, MD).

4.2.2 FPOP labeling of ACTR and CBP

To choose the best condition for ACTR and CBP oxidation, we first performed FPOP experiments on free ACTR and CBP separately. In the FPOP reaction, 5 µM of protein samples in phosphate buffer (10 mM sodium phosphate, 50 mM sodium chloride, pH 7.0) was mixed with glutamine (20 mM) as a radical scavenger limiting the hydroxyl radical lifetime to ~1 µs.[(18)] H_2O_2 was added to the samples just prior to the sample infusion into silica tubing at a flow rate of 19.5 µL/min with an exclusion fraction of 20%[(26)] (the volume not irradiated, existing between plugs of the irradiated protein solution, see Section 1.3.4.2). The KrF excimer laser (GAM Laser Inc., Orlando, FL) frequency was set to 7 Hz and the average power was 45 mJ. The FPOP device setup followed a previously published protocol.[(18)] The conditions were protein sample containing no H_2O_2, 15 mM H_2O_2 but no laser irradiation, 7.5 mM H_2O_2 (half dose) with laser

irradiation, and 15 mM H_2O_2 (full dose) with laser irradiation. After infusing the samples through silica tubing and irradiating with laser pulses, the samples were collected in vials containing 10 nM catalase (to decompose excess H_2O_2) and 40 mM methionine (to scavenge excess secondary oxidants). These samples were initially analyzed as intact proteins by mass spectrometry.

For binding studies, we performed FPOP experiments on free and bound forms of ACTR and CBP. Four sample pools were prepared to represent both ACTR and CBP in free and bound forms. The dissociation equilibrium constant for the ACTR:CBP complex is 34 nM.[22] Samples containing 5 µM ACTR, 5 µM CBP, 5 µM:25 µM (CBP:ACTR), and 25 µM:5 µM (CBP:ACTR) were labeled using the FPOP technique. ACTR:CBP (1:5) will make ACTR 99.8% bound while ACTR:CBP (5:1) will make CBP 99.8% bound. For the free forms of ACTR and CBP, an equal volume phosphate buffer was substituted for the binding partner. All pools were incubated at room temperature for an hour prior to FPOP experiments. 15 mM H_2O_2 (full dose) and 20 mM glutamine were added to each sample prior to labeling with a KrF laser (248nm). The same FPOP protocol was followed as described for the intact proteins. The samples were run in triplicate.

4.2.3 Trypsin digestion

Following irradiation, a 40 µL aliquot of each sample was mixed with 20 µL of tris buffer (100 mM tris-HCl, pH 8.5) and 0.4 µL trypsin (0.1 mg/ml) at an enzyme:protein ratio of 1:20 (w:w). The solution mixtures were incubated at 37 °C overnight. 40 µL aliquots of the digested samples were reconcentrated and cleaned up with NuTip C-18 zip tips and eluted with 30 µL of 60% acetonitrile/40% H_2O/0.1% formic acid. The eluent was then dried and kept at −80 °C until MS analysis.

4.2.4 Mass spectrometry

For the ACTR and CBP intact samples, 5 µL of the samples was mixed with 45 µL of 0.1% formic acid to acidify the solution. The samples were loaded in a silica capillary column, custom-packed with C18 revered phase (Magic, 0.075 mm × 150 mm, 5 µm, 120 Å; Michrom Bioresources, Inc., Auburn, CA, USA) with a custom-pulled tip. The chromatograph was an Ultra 1D+ UPLC (Eksigent, Dublin, CA). The mobile phases were 0.1% formic acid (A) and acetonitrile/0.1% formic acid (B). Samples were eluted using the following gradient: 0% B (0 min), 80% B (60 min), 80% B (70 min), and 0% B (85 min) at 260 nL/min. Mass spectra were collected on a LTQ-FT Ultra controlled by Xcalibur 2.2 software (Thermo-Fisher, San Jose, CA, USA).

For the ACTR and CBP binding studies, the samples were reconstituted in 30 µL of water with 0.1% formic acid. The samples were then analyzed on a LTQ Orbitrap XL (Thermo Fisher, San Jose, CA) operated in data-dependent acquisition mode. An aliquot of 5 µL of each sample was loaded in a silica capillary column, custom-packed with C18 revered phase (Magic, 0.075 mm × 150 mm, 5 µm, 120 Å) with a custom-pulled tip. The chromatograph was an Ultra 1D+ UPLC. Peptides were eluted by a 115 min, 260 nL/min gradient coupled to the nanospray source. The gradient was as follows: 2% B (0 min), 65% B (80 min), 90% B (90 min), 90% B (100 min), 0% B (105 min), and 0% B (115 min). Mass spectra were obtained at high resolving power (50,000, FWHH, for ions of m/z 400). For MS/MS analysis, the six most abundant charge ions with a charge of at least of +2 were subjected to collision-induced diccoiation (CID) using data-dependent scanning.

4.2.5 Data Analysis

For peptide-level data analysis, the raw data MS files were converted to profile mzXML files by using MassMatrix Mass Spec Data File Converter. The mzXML files of all samples were searched against the modified sequence for labeled and unlabeled ACTR and CBP tryptic peptides by using MassMatrix 2.4.2 (Case Western Reserve University, Cleveland, OH).[27] All the known hydroxyl radical side chain reaction products[21, 28] were added to the MassMatrix modification database as variable modifications. Representative MS/MS spectra of modified and unmodified CBP peptides (2109-2117) along with their extracted ion chromatograms (EIC) from the raw data file are shown in Figure 4.1. From the identified modifications on each tryptic peptide, the area under peaks for all modified (AUP_{ox}) and the unmodified (AUP_{unox}) form of each peptide were determined from the raw data using Xcalibur 2.2. The fractions of modification at peptide level were then calculated using equation 1:[29]

$$\text{Fraction of modification of peptide} = \frac{\sum \text{AUP}_{\text{ox}}}{\sum \text{AUP}_{\text{ox}} + \text{AUP}_{\text{unox}}} \qquad (1)$$

The value of fraction of modification is one when no unmodified peptide is detected, while the value is zero when there is no modification. To quantify the changes in solvent accessibility for the ACTR and CBP peptides between the free and bound forms, we calculated protection values (PV) using equation 2:

$$\text{Protection value (PV)} = \frac{\text{Fraction modification in free form}}{\text{Fraction modification in bound form}} \qquad (2)$$

The PV was calculated from the ratio of fraction modification of peptide in the free form to the fraction modification of the same peptide in the bound form.

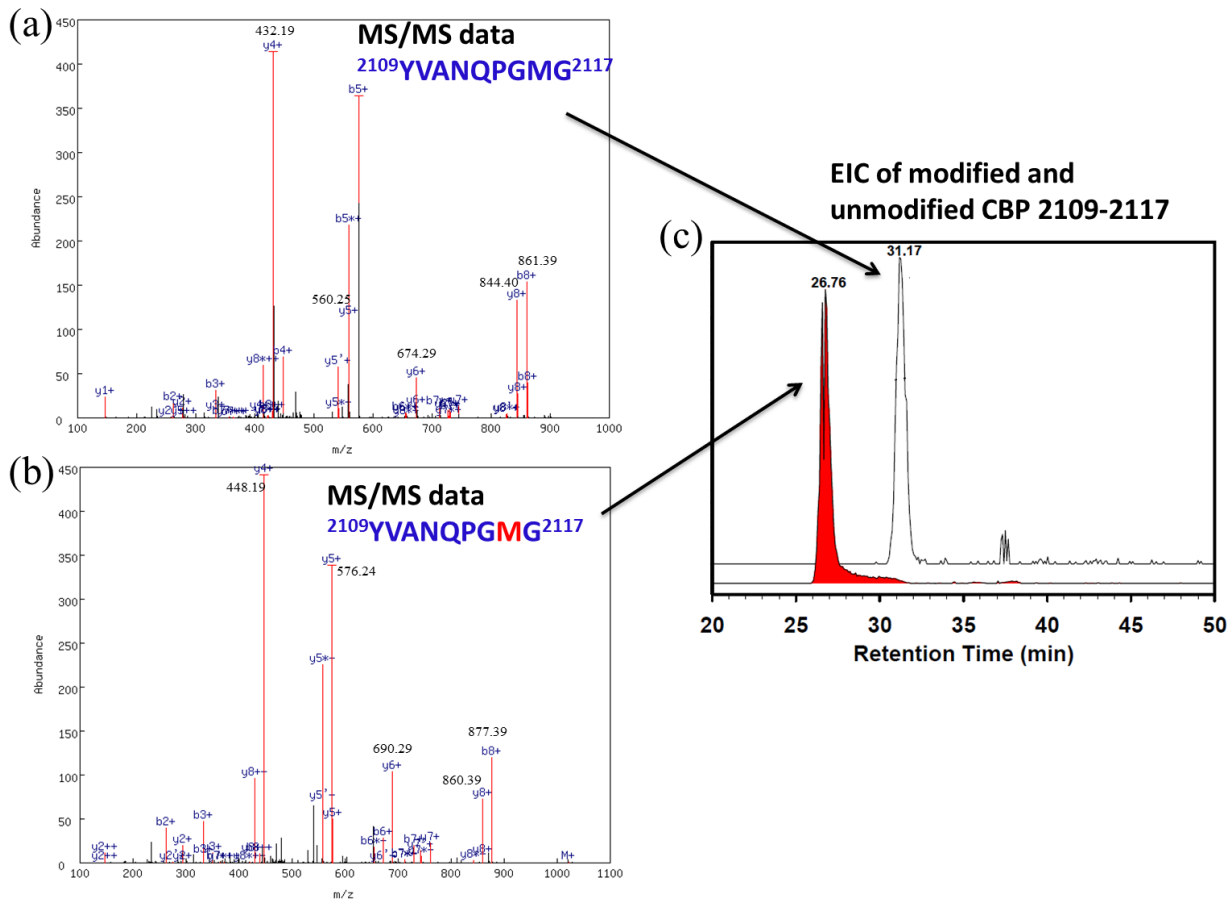


Figure 4.1. FPOP data analysis workflow. Product ion (MS/MS) spectra of CBP peptide 2109-2117 with charge state +2 in the (a) unmodified form and (b) modified form with +16 modification on methionine 2116, colored in red. The area under peak for the peptide was determined from (c) the extracted ion chromatograms of the peptide unmodified (top) and +16 modified (bottom). The fractions of modification can then be determined from the area under peak for the modified (AUP_{ox}) and the unmodified (AUP_{unox}) peaks as described by equation 1. See text for additional details.

4.3 Results and discussion

One of the protein footprinting approaches is FPOP. FPOP can provide structural details about a protein or a protein complex with single-residue resolution.[16, 18, 19] This technique uses short-lived hydroxyl radicals to label the solvent accessible side chains of proteins. The hydroxyl radicals are generated on nanosecond timescales using an excimer laser source. The radical lifetime is controlled to the microsecond timescale using glutamine as a radical scavenger.[18, 20]

Hydroxyl radical has high reactivity that allows irreversible oxidation of 14 of the 20 amino acids. The irreversible oxidation of the side chains makes the FPOP technique adaptable to LC-MS/MS analysis. Probing the solvent accessibility at the residue level makes the FPOP technique useful to study protein folding/unfolding, protein-protein binding, and protein epitope mapping.[26, 29, 30] Over the last decade, the FPOP technique has been widely used to study folded proteins, but there are not many FPOP articles on IDPs. Here we extend the use of FPOP to study IDPs and explore the usefulness of the technique to study this class of proteins. We used ACTR and CBP as model systems for conformational analysis of IDPs. ACTR is a mostly unstructured protein that transiently occupies a folded state as described by NMR and HX-MS.[9, 22, 25, 31, 32] In contrast, CBP is a mostly helical molten globule that transiently occupies an unfolded state as described by NMR and HX-MS.[9, 23, 33]

4.3.1 FPOP condition selection and global analysis of ACTR and CBP

In a typical FPOP experiment, it is necessary to have a mild condition for oxidation because the FPOP approach measures solvent accessibility of a protein. Severe oxidation conditions can alter the native conformational properties of the protein, which in turn leads to very complicated interpretation of the data obtained from FPOP experiments.[16] On the other hand, very mild oxidation conditions can limit the range of local conformations that can be probed.[16] For this purpose, we compared two different H_2O_2 concentrations: half dose (7.5 mM) and full dose (15 mM) for ACTR and CBP oxidation with the control FPOP experiments (no H_2O_2 and no laser irradiation). The mass spectra are shown in Figure 4.2. The most abundant peak for ACTR protein is +7. Therefore, for global data analysis, we used the peak corresponding to +7 charge state from the mass spectra of ACTR oxidation at different conditions. No oxidations were observed for the control experiments, no H_2O_2 added with only

laser irradiation and H_2O_2 added with no laser irradiation, for ACTR. With half dose H_2O_2 for FPOP, we only observed increases in the mass of the ACTR protein in three increments of +16 Da (corresponding to 1103.44, 1105.43, and 1108.07 m/z) as compared to the unmodified protein (1101.16 m/z). With full dose H_2O_2, we observed increases in the mass of the ACTR in five increments of +16 Da (corresponding to 1103.44, 1105.43, 1108.07, 1110.31, and 1112.50 m/z). We also observed low abundance oxygen loss (−14 Da) product for the ACTR using full dose H_2O_2 condition. According to the chemistry of hydroxyl radical reaction with amino acid site chains, oxidation product of −14 Da may correspond to a combination of decarboxylation (−30 Da) and oxygen addition (+16 Da) on an aspartic or glutamic acid side chain [(21)] because ACTR has many aspartic and glutamic acid residues. In the same way, for CBP global analysis, we used the peak corresponding to +7 charge state from the mass spectra of CBP oxidation under different conditions. For the CBP control experiment, no H_2O_2 added with only laser irradiation, two increments of +16 Da were observed. One +16 Da was also observed for the H_2O_2 added with no laser irradiation control condition. CBP has two methionine residues that are highly reactive which makes CBP highly susceptible to oxidation even with no H_2O_2 addition with laser irradiation or H_2O_2 added with no laser irradiation conditions. The oxidation in the control conditions does not invalidate the results because such oxidation is low abundance and cancels out in free and bound comparisons. With half dose H_2O_2 for FPOP, we observed increases in the mass of the CBP in five increments of +16 Da (corresponding to 940.61, 942.91, 945.21, 947.49, and 949.75 m/z) as compared to the unmodified protein (938.37 m/z). With full dose H_2O_2 for FPOP, we observed increases in the mass of the CBP in seven increments of +16 Da (corresponding to 940.61, 942.91, 945.21, 947.49, 949.75, 952.08, and 954.36 m/z). Using full dose H_2O_2 condition for CBP, we also observed (−16 Da) product, which may correspond to

oxygen loss on an aspartic acid side chain because CBP has one aspartic acid residue. Although both H_2O_2 doses, half and full, provided relatively good conditions for the ACTR and CBP oxidation, we selected full H_2O_2 dose concentration as a base to perform FPOP experiments at the peptide level analysis. The reason for using full dose is because the extent of modification for all the ACTR and CBP peptides in the free state were at least 0.05, in which the chances for observing the actual differences in two protein states is very likely. Thus, with the full H_2O_2 dose, a wide range of local conformations of ACTR and CBP can be probed upon binding to each other.

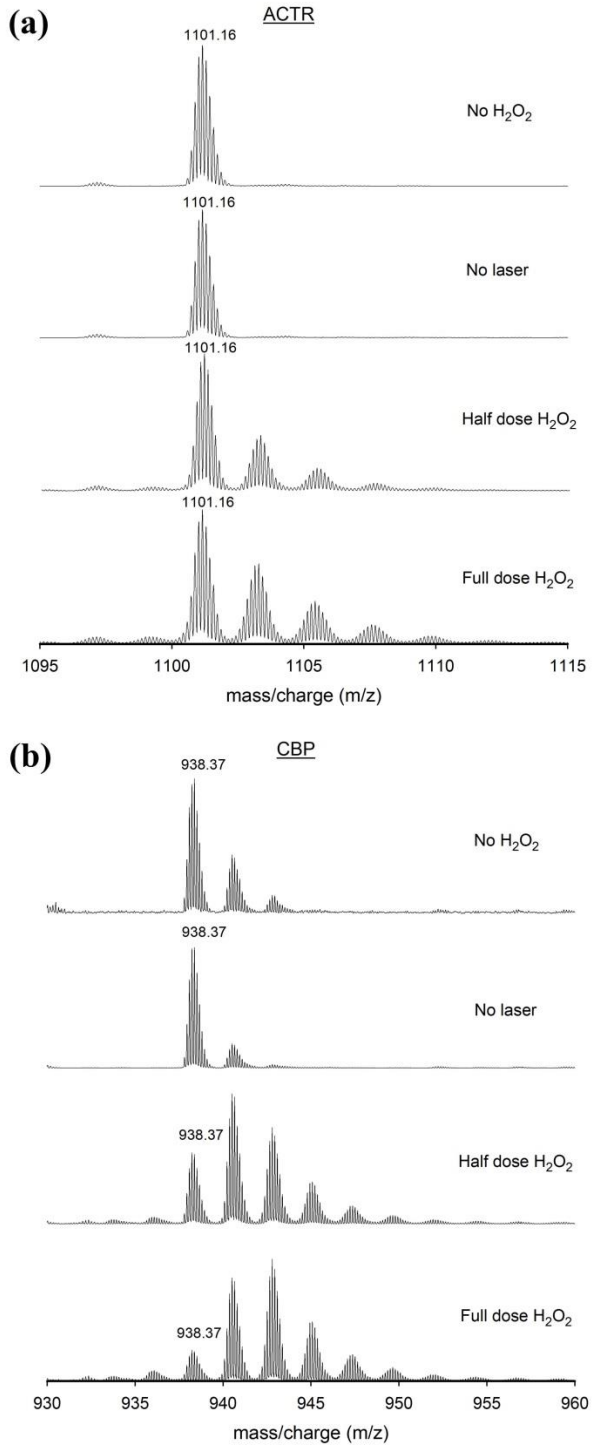


Figure 4.2. Global level analysis of (a) ACTR and (b) CBP after FPOP labeling for the protein charge state of +7. The FPOP experiments were conducted under four different conditions: no H_2O_2 added to the protein solution with laser irradiation, added 15 mM H_2O_2 with no laser irradiation, 7.5 mM H_2O_2 (half dose) with laser irradiation, and 15 mM H_2O_2 (full dose) with laser irradiation.

Although we mainly performed the global FPOP experiments to select the best oxidation condition, the data can also be used to compare the extent of oxidation of the two IDPs under the same oxidation conditions. For example, using half dose H_2O_2 conditions for the FPOP, we observed three increments of +16 Da for ACTR with low abundance of modifications as compared to five increments of +16 Da for CBP with high abundance of modifications. We also observed that CBP is more oxidized than ACTR when we compare the full dose H_2O_2 experiment for ACTR and CBP. According to global FPOP data, CBP appears to be less protected than ACTR. The data are inconsistent with what is expected for ACTR and CBP. ACTR is a nearly unstructured protein that transiently occupies a folded state as described by NMR and HX-MS.[9, 22, 31] However, CBP is a mostly helical molten globule that transiently occupies an unfolded state as described by NMR and HX-MS.[9, 23, 33] The comparison of global FPOP data between ACTR and CBP is not very useful. The data cannot provide the structure contents when we compare two IDPs to each other. The reason is because the oxidation of a given protein or peptide does not only depend on the solvent exposure of the amino acid side chains but also depends on the reactivity of the side chains. CBP has more reactive side chains than ACTR. Two highly reactive methionine residues are present in the sequence of CBP with rate constant of 8.5×10^9 M^{-1} s^{-1} (see Table 4.1). Therefore, the FPOP data are not useful for comparing the structure content of two different proteins with each other. This is why most FPOP applications are comparisons of two states of the same protein.

4.3.2 Peptide-level FPOP data analysis of ACTR and CBP

We performed FPOP labeling of ACTR with full dose H_2O_2 in the absence and presence of five-fold molar excess of CBP. In the same way, we performed FPOP labeling of CBP with full dose H_2O_2 in the absence and presence of five-fold molar excess of ACTR. For ACTR,

tryptic digestion yielded four peptides: ACTR 1023-1031, 1032-1051, 1052-1074, and 1075-1091. The ACTR sequence coverage was 97%. We observed changes in the extent of modification in only two C-terminal peptides when we compared the free form of ACTR with the bound form of ACTR. For CBP, tryptic digestion also yielded four peptides: CBP 2062-2073, 2077-2092, 2093-2103, and 2109-2117. The CBP sequence coverage was 81%. Figure 4.3 shows the fraction of modification of the ACTR and CBP peptides in their free and bound forms. We observed changes in the extent of modification in three peptides when we compared the free form CBP with bound form of CBP. The blue colored bars in the Figure 4.3 denote the fraction of modification of the peptides obtained from the free forms of the proteins (*e.g.* 1 ACTR). The red colored bars denote the fraction of modification of the peptides obtained from bound forms of the protein in the presence of five-fold molar excess of the partner protein (*e.g.* 1 ACTR: 5 CBP). We also included the fraction of modification data for the peptides obtained from bound forms of the protein but in the presence of five-fold molar excess of the target protein (*e.g.* 5 ACTR: 1 CBP) (diamond patterned bars in Figure 4.3). The estimated fraction of modification data for (5 ACTR: 1 CBP) (line patterned bars in Figure 4.3) were also included and discussed below.

We compared the fractions of modification for the free form of ACTR peptides with each other (blue colored bars in Figure 4.3a). For example, in the N-terminal peptide of ACTR 1023-1031, in the absence of CBP, we observed the least modification with the average value of 0.05. According to the NMR and HX data, the N-terminal region of ACTR is unstructured in the free form,[(9, 22)] thus we would expect a large fraction of modification. However, we observed the opposite result. The low fraction of modification for this region is because this region is enriched in amino acid side chains with low reactivity such as glycine, threonine, glutamine, and

asparagine (see Table 4.1). On the other hand, for the middle region of ACTR, 1052-1074, we observed the highest degree of modification with an average value of 0.8. Again according to the NMR and HX data, this region of the free form of ACTR has some residual structure with more structuring than the N-terminal region, thus we expected to observe less modification than in the N-terminal region of the ACTR. The middle region of ACTR contains relatively more reactive side chains. Therefore, the fraction of modification of the free form of ACTR alone is not useful for the structural interpretation of the data. Like the global protein analysis, the fractions of modification for the free form of ACTR peptides do not report about the relative structure contents.

We also compared the fractions of modification for the free form of CBP peptides with each other (blue colored bars in Figure 4.3b). For example, for the C-terminal peptide of CBP 2109-2117 in the absence of ACTR, we observed the highest amount of modification with the average value of 0.87. According to the NMR and HX data, the N-terminal region of CBP is unstructured in the free form[9, 22] and we would expect more modification in the N-terminal region than in the C-terminal region. However, we observe the opposite result because the C-terminal region of CBP contains highly reactive amino acid side chains such as methionine. In contrast, we observed less modification for the N-terminal peptides because of the absence of highly reactive amino acid side chains. Therefore, the fraction of modification of the free form of CBP alone is not useful for the structural interpretation of the data.

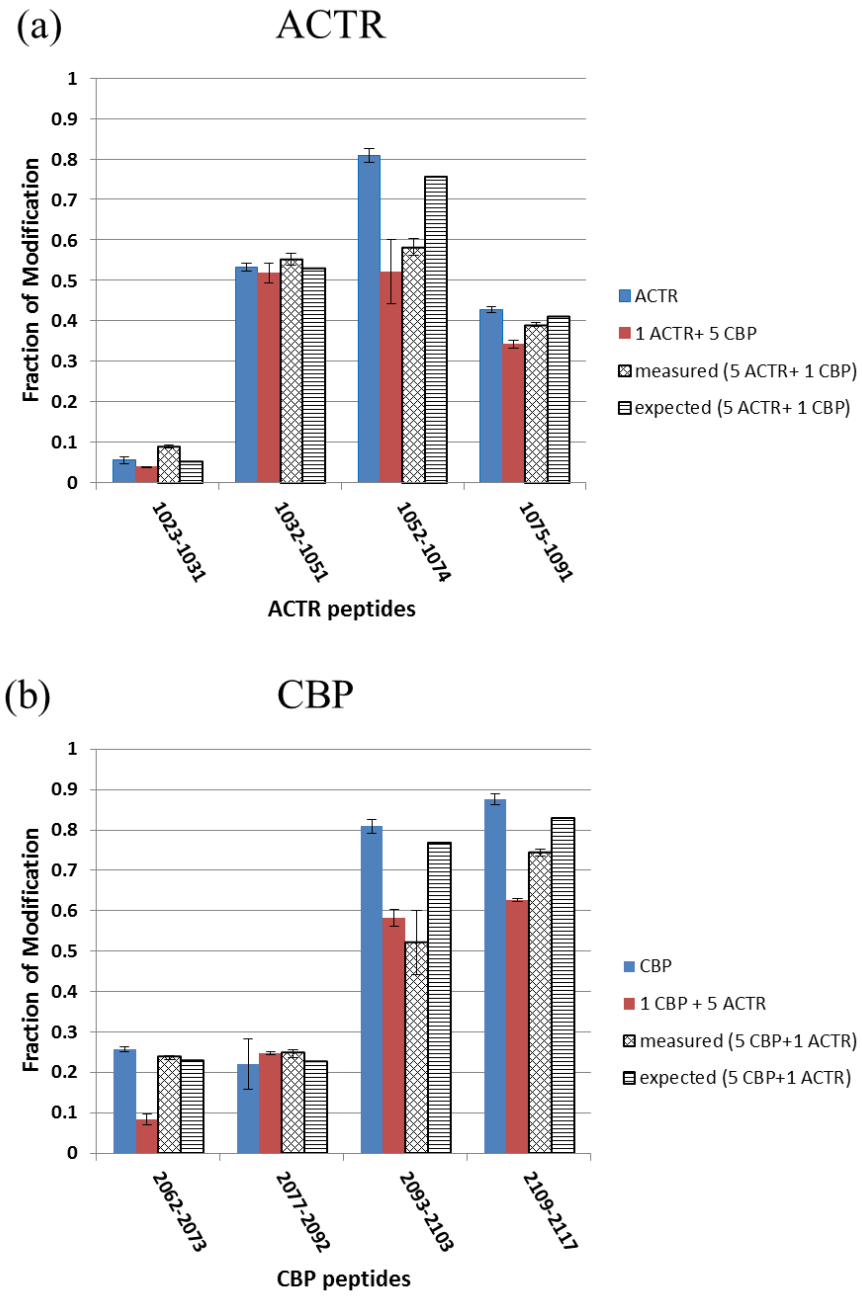


Figure 4.3. Fractions of modification of (a) ACTR and (b) CBP at the peptide level. The fractions of modification were calculated as described in the Materials and Methods section. The fractions of modification were determined for ACTR free form (blue bars), ACTR bound to CBP at 1:5 molar ratio (red bars), ACTR bound to CBP at 5:1 molar ratio (diamond pattern bars), and estimated for ACTR bound to CBP at 5:1 molar ratio (line pattern bars). The fractions of modification were determined for CBP free form (blue bars), CBP bound to ACTR at 1:5 molar ratio (red bars), CBP bound to ACTR at 5:1 molar ratio (diamond pattern bars), and estimated for CBP bound to ACTR at 5:1 molar ratio (line pattern bars). The error bars denote standard deviation from triplicate measurements.

To test the reliability of FPOP data, we compared measured with estimated extent of modification of peptides obtained from samples with high protein concentrations. This comparison would provide information about whether the excess concentration of a protein partner has an effect on the dosage of hydroxyl radical delivered to the target protein. For ACTR, there are two experiments where the extent of modification of free ACTR peptide is measured: [5 μM ACTR, blue bars in Figure 4.3a] and [25 μM ACTR + 5 μM CBP, diamond patterned bars in Figure 4.3a]. There are also two experiments where the extent of modification of bound ACTR peptide is measured: [25 μM ACTR + μM 5 CBP, diamond pattern bars in Figure 4.3a] and [5 μM ACTR + 25 μM CBP, red bars in Figure 4.3a]. In the first experiment, [5 μM ACTR], ACTR is 100% free. Based on the equilibrium calculation and the dissociation equilibrium constant for ACTR:CBP complex, ACTR is approximately 82% free and 18% bound in the second experiment, [25 μM ACTR + 5 μM CBP]. The observed oxidation would then be a weighted average of 82% free ACTR and 18% bound ACTR if the extent of oxidation did not depend on the presence of CBP. This approximation will only work if the lifetime of the hydroxyl radicals is much shorter than the lifetime of the ACTR:CBP complex. If the hydroxyl radical lifetime, controlled by hydrogen peroxide concentration and glutamine as radical scavenger, is 1 μs, as claimed in Hambly and Gross,[18] then this is probably a reasonable assumption. If the hydroxyl radical lifetime is longer than microsecond timeframe, protein complex molecules, depending on the dissociation constant, can change from bound to free (or reverse) during the lifetime of the radicals, which in turn leading to difficulties in the interpretation of the FPOP data. In the fourth experiment, [5 μM ACTR + 25 μM CBP], ACTR is approximately 99.8% bound. Therefore, the extent of modification for each peptide in the middle experiment, [25 μM ACTR + 5 μM CBP], was estimated (line patterned bars in Figure

4.3a) based on how much modification observed in the other two experiments. In the same way, the extent of modification for each peptide of CBP in the middle experiment, [25 μM CBP + 25 μM ACTR], was estimated (line patterned bars in Figure 4.3b). As shown in the Figure 4.3a, there are deviations between the extent of oxidation of measured (diamond patterned bars) from the estimated (line patterned bars) for all ACTR peptides. We observed less measured oxidation for peptides (1052-1074 and 1075-1091) and more measured oxidation for peptides (1023-1031 and 1032-1051) than the estimated values. As shown in the Figure 4.3b, there are deviations of the extent of oxidation of measured from the estimated for three CBP peptides. We observed less measured oxidation for peptides (2093-2103 and 2109-2117) and more measured oxidation for peptides (2077-2092) than the estimated values. We also observed that for ACTR and CBP regions that become heavily modified, the measured amount of modification is less than expected. Whereas in less reactive regions, the measured amount of modification is about the expected amount. On the basis of these deviations of measured from estimated oxidation, the data indicate that the extra protein sometimes interferes with the hydroxyl radical labeling. One of the reasons might be that extra protein consumes the hydroxyl radicals. The other reason might be that the hydroxyl radical lifetime is longer than 1 μs as claimed before. In a very recent paper,[(34)] Vahidi and Konermann showed that the oxidation in a typical FPOP experiment might extend over tens of millisecond instead of 1 μs. This new millisecond time range for oxidation is caused by secondary radicals generated by glutamine scavenger in the FPOP solution. The deviations of measured from estimated oxidations could also be explained by both reasons: hydroxyl radical consumption by extra protein and extended oxidation over longer timescale than 1 μs.

4.3.3 Regions of decreased solvent accessibility of ACTR upon binding to CBP

The extent of oxidation for ACTR can be compared in the absence and presence of CBP. Figure 4.3a shows the fraction of modification of each ACTR peptide in the free and bound states. The data shown in Figure 4.3a were converted to protection values (PVs) (see equation 2) and the PVs were mapped onto the amino acid sequence of ACTR, as shown in Figure 4.4. The ACTR peptides, 1023-1031 and 1032-1051, showed no or slight differences in fractions of modification between their free and bound forms and the protection values were in the range 0.9 to 1.2 (see Figure 4.3a and 4.4), suggesting that CBP has no effect on the N-terminal segments of ACTR. These peptides are located in the N-terminal region of ACTR which is an important region to validate the FPOP approach because NMR and HX data revealed that this region is unstructured before and after CBP binding.[9, 31] In contrast, we observed a substantial increase in protection from oxidation in ACTR 1052-1074 and 1075-109 upon CBP binding. These peptides cover the middle and C-terminal regions of ACTR (see Figure 4.3a and 4.4). The PVs for these two peptides were in the range of 1.2-2.5, and the PVs are the highest degree of protection observed in ACTR. These ACTR peptides are located in the regions that form α-helices in the complex as revealed by NMR. In general, the FPOP data, as protection values, are consistent with the NMR and HX data presented for all regions of bound ACTR.

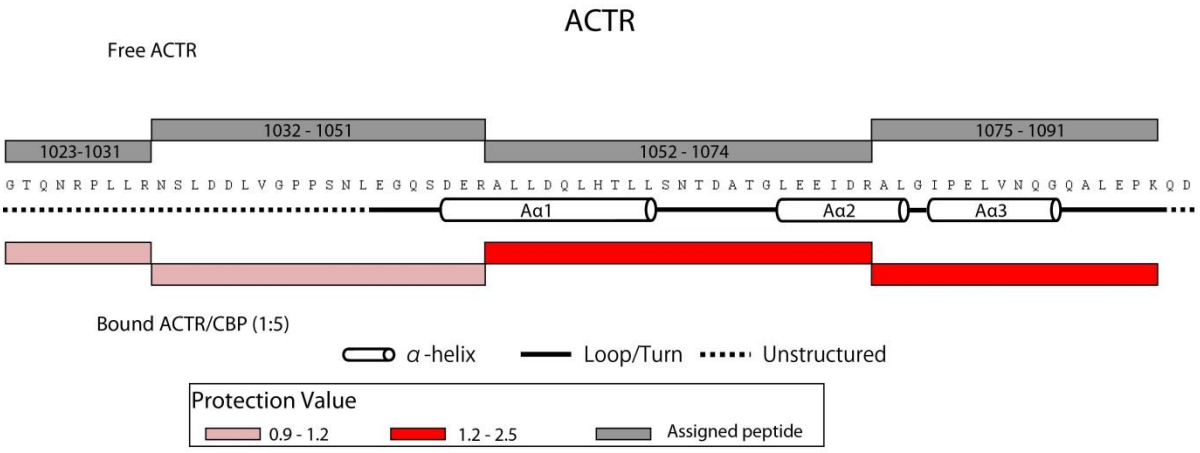


Figure 4.4. Protection values mapped onto the primary and secondary structure of the bound form of ACTR for each peptide. The ACTR tryptic peptides shown in gray cover 97% of the sequence. The secondary structural elements from the ACTR/CBP complex [(22)] are also indicated.

4.3.4 Regions of decreased solvent accessibility of CBP upon binding to ACTR

The extent of oxidation for CBP can be compared in the absence and presence of ACTR. The fraction of modification data from Figure 4.3b were converted to PVs and the PVs were then mapped on the amino acid sequence of CBP, as presented in Figure 4.5. The results for most of the CBP peptides in the bound form showed increased protection from oxidation relative to the free form. For example, binding of ACTR to CBP caused substantial protection from oxidation of peptide CBP 2062-2073. We observed a 3-fold increase in protection from oxidation for this peptide located in the N-terminal region of CBP. This is the highest amount of protection observed across the sequence of CBP. The C-terminal CBP peptides, 2093-2103 and 2109-2117, were also protected from oxidation as shown in Figure 4.5, but the change in the amount of protection is not as much in the N-terminal segment of CBP. The PVs for the C-terminal peptides were in the range of 1.2-2.5. The data suggest that the N- and C-terminal regions of CBP become more tightly folded upon binding to ACTR. In general, we observed good agreement between the FPOP data for the N-terminal and C-terminal regions and the NMR and

HX data.[9, 31] On the other hand, there is one peptide, CBP 2077-2092, in the middle region of CBP that showed no difference in the protection between the free and bound forms of CBP. This peptide covers a turn between α1 and α2-helices that is not affected by binding. We observed the lowest amount of PV for this peptide across all regions of CBP. The data for this region do not agree well with NMR and HX data. The reason is because this peptide is enriched in residues that have side chains with low reactivity with hydroxyl radicals: asparagine, serine, lysine, and glutamine. The weakly reactive side chains may decrease the ability to measure the degree of changes in the protection for the CBP 2077-2092.

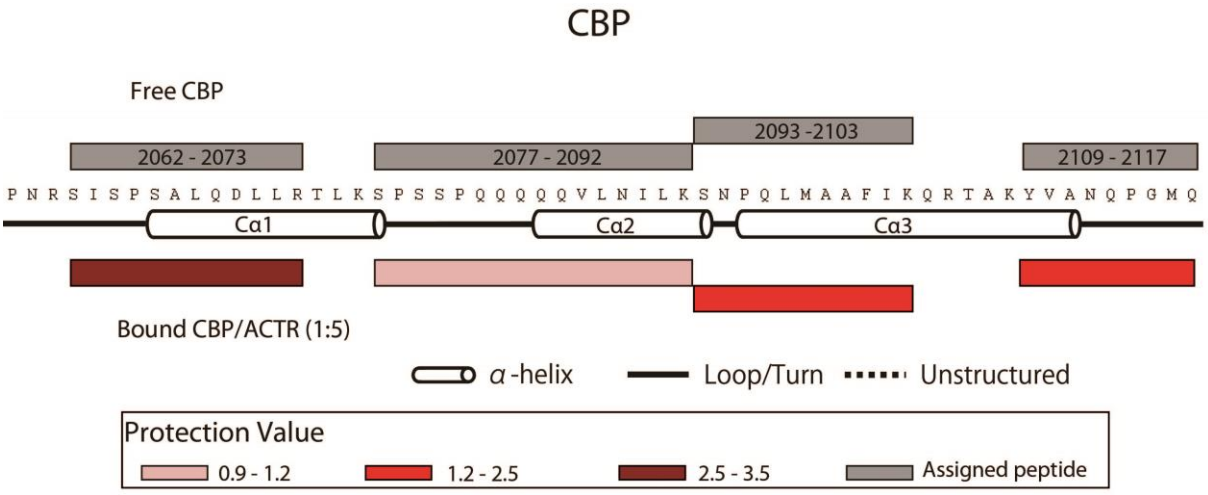


Figure 4.5. Protections values mapped onto the primary and secondary structure of the bound form of CBP for each peptide. The CBP tryptic peptides shown in gray cover 81% of the sequence. The secondary structural elements from the ACTR/CBP complex [22] are also indicated.

The FPOP data for ACTR and CBP illustrate the potential usefulness of FPOP analysis to characterize and localize regions of binding induced folding of IDPs. However, FPOP may not work for all kinds of IDPs, and especially IDPs that are highly enriched in amino acids that are weakly reactive with hydroxyl radicals. Amino acids predicted to be enriched in IDPs, including alanine, glycine, glutamine, serine, and glutamic acid, generally have smaller rate constants (see

Table 4.1).[21, 35, 36] Whereas many amino acids that are enriched in ordered proteins, including tryptophan, phenylalanine, cysteine, and isoleucine, generally react more quickly with hydroxyl radicals relative to other amino acids (see Table 4.1).[21, 35] Unlike HX, the application of the FPOP method to interpret the structure content of the free forms of a protein is not very useful. The reason is because in the FPOP method, the measured extent of oxidation of a given peptide largely depends not only on the solvent exposure of the amino acid side chains in the protein's structure but also the reactivity of amino acid side chains with hydroxyl radicals. When a protein contains amino acid side chains with low reactivity, then the chances of obtaining structural informational are less.

In order to confidently characterize conformational change of a protein under various conditions, the FPOP method can be used as a stand-alone method or it can work as a complementary technique to other techniques such as HX that can probe protein structure. The FPOP method would be beneficial to achieve high spatial resolution for protein conformation and dynamic studies, but if a protein is enriched in amino acid side chains with low reactivity, FPOP cannot be very useful. There are also problems associated with FPOP experiments. One of the problems is reproducibility. The reproducibility of FPOP experiments highly depends on the ability to deliver a defined concentration of hydroxyl radicals to the protein of interest. Delivering a defined concentration of radicals not only depends on the amount of radicals generated and presence of radical scavenger, but also depends on some other factors: the concentration of the proteins in the FPOP solution, buffer components, and even the geometry of each FPOP setup that may alter the amount of radicals produced and delivered to the protein target. In such cases, to measure a defined concentration of hydroxyl radicals generated and delivered to the protein target and reliably measure oxidation of the protein, a radical dosimeter

is required in FPOP experiment. Recently, adenine as a radical dosimeter has been reported to measure the effective hydroxyl radicals delivered to the protein of interest.[37] The reason for selecting adenine as a radical dosimeter is due to its quantitative relationship between UV spectrophotometric response, hydroxyl radical dose, and protein oxidation levels. One other problem associated with FPOP is the radical lifetime available for protein labeling. The radical lifetime subject is still under debate.[18, 26, 34] As mentioned before, the hydroxyl radical lifetime is claimed to be ~1 μs, and in such time frame protein labeling will be complete before conformational changes can take place.[18, 26] A recent article suggests that the oxidation can extend over milliseconds through secondary radical products produced from reaction of hydroxyl radicals with glutamine scavenger in FPOP solution.[34] If the new time frame for radical lifetime is correct, the oxidation might impact the native conformation of a target protein. In cases of long-lived radicals, the problems could also arise if the oxidation is not much faster than the on/off rate for the protein complex.

4.4 References

1. Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions?, *Proteins 41*, 415-427.

2. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) Intrinsic disorder and protein function, *Biochemistry 41*, 6573-6582.

3. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nature reviews. Molecular cell biology 6*, 197-208.

4. Tantos, A., Han, K.-H., and Tompa, P. (2012) Intrinsic disorder in cell signaling and gene transcription, *Molecular and cellular endocrinology 348*, 457-465.

5. Radivojac, P., Iakoucheva, L. M., Oldfield, C. J., Obradovic, Z., Uversky, V. N., and Dunker, A. K. (2007) Intrinsic disorder and functional proteomics, *Biophysical journal 92*, 1439-1456.

6. Vasilescu, J., Guo, X., and Kast, J. (2004) Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry, *Proteomics 4*, 3845-3854.

7. Wales, T. E., and Engen, J. R. (2006) Hydrogen exchange mass spectrometry for the analysis of protein dynamics, *Mass spectrometry reviews 25*, 158-170.

8. Iacob, R. E., Murphy, J. P., 3rd, and Engen, J. R. (2008) Ion mobility adds an additional dimension to mass spectrometric analysis of solution-phase hydrogen/deuterium exchange, *Rapid communications in mass spectrometry : RCM 22*, 2898-2904.

9. Keppel, T. R., Howard, B. A., and Weis, D. D. (2011) Mapping Unstructured Regions and Synergistic Folding in Intrinsically Disordered Proteins with Amide H/D Exchange Mass Spectrometry, *Biochemistry 50*, 8722-8732.

10. Rob, T., Liuni, P., Gill, P. K., Zhu, S., Balachandran, N., Berti, P. J., and Wilson, D. J. (2012) Measuring dynamics in weakly structured regions of proteins using microfluidics-enabled subsecond H/D exchange mass spectrometry, *Analytical chemistry 84*, 3771-3779.

11. Goswami, D., Devarakonda, S., Chalmers, M. J., Pascal, B. D., Spiegelman, B. M., and Griffin, P. R. (2013) Time window expansion for HDX analysis of an intrinsically disordered protein, *Journal of the American society for mass spectrometry 24*, 1584-1592.

12. Keppel, T. R., and Weis, D. D. (2013) Analysis of disordered proteins using a simple apparatus for millisecond quench-flow H/D exchange, *Analytical chemistry 85*, 5161-5168.

13. Yang, J., Garrod, S. M., Deal, M. S., Anand, G. S., Woods, V. L., Jr., and Taylor, S. (2005) Allosteric network of cAMP-dependent protein kinase revealed by mutation of Tyr204 in the P+1 loop, *Journal of molecular biology 346*, 191-201.

14. Kheterpal, I., and Wetzel, R. (2006) Hydrogen/deuterium exchange mass spectrometry--a window into amyloid structure, *Accounts of chemical research 39*, 584-593.

15. Chalmers, M. J., Busby, S. A., Pascal, B. D., He, Y., Hendrickson, C. L., Marshall, A. G., and Griffin, P. R. (2006) Probing protein ligand interactions by automated hydrogen/deuterium exchange mass spectrometry, *Analytical chemistry 78*, 1005-1014.

16. Fitzgerald, M. C., and West, G. M. (2009) Painting proteins with covalent labels: what's in the picture?, *Journal of the American Society for Mass Spectrometry 20*, 1193-1206.

17. Konermann, L., Pan, J., and Liu, Y. H. (2011) Hydrogen exchange mass spectrometry for studying protein structure and dynamics, *Chemical society reviews 40*, 1224-1234.

18. Hambly, D. M., and Gross, M. L. (2005) Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale, *Journal of the American society for mass spectrometry 16*, 2057-2063.

19. Hambly, D., and Gross, M. (2007) Laser flash photochemical oxidation to locate heme binding and conformational changes in myoglobin, *International journal of mass spectrometry 259*, 124-129.

20. Hambly, D. M., and Gross, M. L. (2008) Microsecond time-scale hydroxyl radical profiling of solvent-accessible protein residues, In *Comprehensive analytical chemistry* (Julian, P. W., Ed.), pp 151-177, Elsevier.

21. Xu, G., and Chance, M. R. (2007) Hydroxyl radical-mediated modification of proteins as probes for structural proteomics, *Chemical reviews 107*, 3514-3543.

22. Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J., Evans, R. M., and Wright, P. E. (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators, *Nature 415*, 549-553.

23. Ebert, M. O., Bae, S. H., Dyson, H. J., and Wright, P. E. (2008) NMR relaxation study of the complex formed between CBP and the activation domain of the nuclear hormone receptor coactivator ACTR, *Biochemistry 47*, 1299-1308.

24. Kjaergaard, M., Teilum, K., and Poulsen, F. M. (2010) Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP, *Proceedings of the national academy of sciences of the United States of America 107*, 12535-12540.

25. Keppel, T. R., and Weis, D. D. (2015) Mapping residual structure in intrinsically disordered proteins at residue resolution using millisecond hydrogen/deuterium exchange and residue averaging, *Journal of the American society for mass spectrometry 26*, 547-554.

26. Gau, B. C., Sharp, J. S., Rempel, D. L., and Gross, M. L. (2009) Fast photochemical oxidation of protein footprints faster than protein unfolding, *Analytical chemistry 81*, 6563-6571.

27. Xu, H., Yang, L., and Freitas, M. A. (2008) A robust linear regression based algorithm for automated evaluation of peptide identifications from shotgun proteomics by use of reversed-phase liquid chromatography retention time, *BMC bioinformatics 9*, 347.

28. Xu, G., and Chance, M. R. (2004) Radiolytic modification of acidic amino acid residues in peptides: probes for examining protein-protein interactions, *Analytical chemistry 76*, 1213-1221.

29. Zhang, H., Gau, B. C., Jones, L. M., Vidavsky, I., and Gross, M. L. (2011) Fast photochemical oxidation of proteins for comparing structures of protein-ligand complexes: the calmodulin-peptide model system, *Analytical chemistry 83*, 311-318.

30. Jones, L. M., J, B. S., J, A. C., and Gross, M. L. (2011) Fast photochemical oxidation of proteins for epitope mapping, *Analytical chemistry 83*, 7657-7661.

31. Kjaergaard, M., Norholm, A. B., Hendus-Altenburger, R., Pedersen, S. F., Poulsen, F. M., and Kragelund, B. B. (2010) Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II?, *Protein science : a publication of the protein society 19*, 1555-1564.

32. Kjaergaard, M., Iešmantavičius, V., and Poulsen, F. M. (2011) The interplay between transient α-helix formation and side chain rotamer distributions in disordered proteins probed by methyl chemical shifts, *Protein science : a publication of the protein society 20*, 2023-2034.

33. Lin, C. H., Hare, B. J., Wagner, G., Harrison, S. C., Maniatis, T., and Fraenkel, E. (2001) A Small Domain of CBP/p300 Binds Diverse Proteins: Solution Structure and Functional Studies, *Molecular cell 8*, 581-590.

34. Vahidi, S., and Konermann, L. (2016) Probing the Time Scale of FPOP (Fast Photochemical Oxidation of Proteins): Radical Reactions Extend Over Tens of Milliseconds, *Journal of the American society for mass spectrometry 27*, 1156-1164.

35. Williams, R. M., Obradovi, Z., Mathura, V., Braun, W., Garner, E. C., Young, J., Takayama, S., Brown, C. J., and Dunker, A. K. (2001) The protein non-folding problem: amino acid determinants of intrinsic order and disorder, *Pacific symposium on biocomputing*, 89-100.

36. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein, *Proteins 42*, 38-48.

37. Xie, B., and Sharp, J. S. (2015) Hydroxyl Radical Dosimetry for High Flux Hydroxyl Radical Protein Footprinting Applications Using a Simple Optical Detection Method, *Analytical chemistry 87*, 10719-10723.

Chapter five

Conclusion and future directions

5.1 Overview

Intrinsically disordered proteins (IDPs) are a class of proteins that exhibit a high degree of conformational flexibility and lack a structured conformation under physiological conditions, but are still functional.[1-3] IDP functionality arises in a different manner as compared to the classical structure-function model of ordered proteins. There has been a growing interest in IDPs due to their important roles in many biological processes and functions. Highly dynamic conformations in IDPs mediate protein-protein interaction through the phenomenon of coupled folding and binding.[4-6] Two extreme mechanisms have been proposed to describe this phenomenon: conformational selection and induced folding.[7, 8] The differentiation of these two mechanisms is an ongoing challenge. More detailed information is required to understand the complexity of mechanisms of protein-protein interaction involving disordered proteins. Also, localized information about the presence or absence of transiently folded conformers of IDPs in the free state can be very important to provide support for one proposed mechanism over the other. Since IDPs are unfolded or partially unfolded and are very flexible, they have a high solvent accessible surface. These properties of IDPs represent a challenge for biophysical techniques to analyze this class of proteins. Over the past two decades, mass spectrometry (MS) has become a powerful technique for biophysical characterization of proteins due to its high sensitivity and variety of choices for sample preparation and instrumentation.[9-11] This dissertation presented examples of the application of two important mass spectrometry-based approaches: hydrogen exchange (HX) and fast photochemical oxidation of proteins (FPOP) to study disordered protein systems. The data illustrate the usefulness of the mass spectrometry-based approaches to characterize and localize regions of protein-protein interactions involving intrinsically disordered proteins.

5.2 Chapter summaries and future directions

5.2.1 Chapter 2

HX-MS was applied to better understand the mechanism of calcineurin activation. Calcineurin is a heterodimeric phosphatase that plays essential roles in cellular processes. Previous work has established that at high calcium concentration, calmodulin binds calcium ions, resulting in calmodulin binding to the intrinsically disordered regulatory domain of calcineurin.[12-14] Calmodulin binding causes release of the autoinhibitory domain from the active site, activating calcineurin. Our results with full-length calcineurin demonstrated that the regulatory domain is unstructured in the absence of calmodulin, while it folds upon binding to calmodulin. This result confirms previous work on the isolated regulatory-autoinhibitory domain construct.[12] Additionally, we have observed calmodulin-induced changes in peptides located in other domains of calcineurin. Finally, and most surprisingly, we found no changes in the structuring of the calcineurin autoinhibitory domain upon calmodulin binding. We presented results from all regions of the calcineurin to describe the mechanism of calcineurin activation. We also proposed a new model of calcineurin activation upon calmodulin binding.

The HX data presented in Chapter 2 described only two states of calcineurin at high calcium concentrations: free and bound to calmodulin. There is limited information about calcineurin conformation at no or low calcium concentrations in the absence of calmodulin. The available information about calcineurin in this state is based on limited proteolysis and some other low resolution biochemical studies. According to previous studies, in the absence of calcium and calmodulin, the calmodulin-binding domain and autoinhibitory domain of the calcineurin regulatory domain are protected against proteolytic attack, indicating these two regions of the regulatory domain may remain packed against the calcineurin catalytic domain.[14]

To investigate the complex multistep regulatory mechanism of calcineurin activation and have a complete story of the activation, future studies on calcineurin in the absence of calcium in solution will be required. It will be interesting to perform HX-MS experiments to study calcineurin in the absence of calcium state. For this purpose, HX-MS can be performed on calcineurin in the absence and presence of EDTA as a calcium free state.

According to the HX data presented in Chapter 2, we also observed calcineurin conformational changes in the C-terminal region of calcineurin B chain and the region immediately N-terminal to the autoinhibitory domain of calcineurin A chain. We suggested that the C-terminal region of calcineurin B and the region immediately N-terminal to the autoinhibitory domain may interact with calmodulin to anchor calmodulin to the binding site. To confirm that there is an interaction between the C-terminal region of calcineurin B and calmodulin, we also need to study calmodulin conformational changes at high calcium concentration in the presence and absence of calcineurin using HX-MS. The HX-MS experiment will be challenging. The reason is because calmodulin itself undergoes conformational change when it bound to calmodulin binding domain of calcineurin.[12, 15] Thus, distinguishing between HX protection at the calcineurin-calmodulin interface from binding-induced folding of calmodulin will not be easy. However, a control HX experiment will be required in such cases. HX-MS on calmodulin in the presence of calmodulin binding domain peptide can be used as a control experiment. The HX data of the control experiment will then be compared with HX data for the calmodulin in the presence of calcineurin to be able to observe if there is interaction between the C-terminal region of calcinerin B chain and calmodulin.

5.2.2 Chapter 3

The degree of structure measurement in IDPs can provide important information about the mechanisms by which IDPs undergo coupled folding and binding. Different approaches to quantify the degree of structure in IDPs using millisecond HX were explored. A quench-flow device, built in-house,[16] for HX labeling on the millisecond timescale was employed. It is essentially impossible to determine the degree of structure without having an accurate unprotected reference state. The interaction domains of the activator for thyroid and retinoid receptors (ACTR) and the CREB binding protein (CBP) were used as model IDPs to explore the best approach to produce an unprotected reference state for millisecond HX. ACTR is a near-random coil IDP that has some residual helicity while CBP is a molten globular IDP that transiently becomes unstructured as revealed by NMR and HX-MS.[17-19] The approaches explored to obtain an unprotected reference state in HX experiments were chemical exchange calculations, addition of denaturing agents, and millisecond HX labeling of peptic peptides obtained from the IDP. We found that peptic reference peptides can be used as an accurate unprotected reference state for determining the degree of the structure.

The approach for determining the degree of structure discussed in Chapter 3 may not work for all types of proteins. If a peptide can form structure even after protein digestion, then these peptides cannot act as an unprotected reference peptide. This approach could be a problem for globular proteins that have a much stronger tendency to fold. This could also be a problem for IDPs that contain regions that have a strong tendency to fold. It will be interesting to develop a method to account for such issues. One possible method is by using peptic peptide in a low denaturant concentration as an unprotected reference state for the degree of structure quantification at high resolution.

Since we successfully applied the approach discussed in Chapter 3 to quantify the degree of structure of ACTR and CBP as two model IDPs, it will be interesting to perform similar studies to quantify the degree of structure of calcineurin in the absence of calmodulin. This experiment will provide structural and flexibility information about calcineurin with a very simple and easy to apply experimental setup.

5.2.3 Chapter 4

Due to its fast labeling timeframe, FPOP might reveal states of IDPs that are undetectable by HX. The application of the FPOP technique for characterizing IDPs was evaluated. To explore the applicability of this technique for studying IDPs, ACTR and CBP as model systems that co-fold upon binding were used. The FPOP technique was utilized to study ACTR and CBP in their free and bound forms. The data showed that FPOP can provide useful information to compare two states of IDPs. The usefulness and limitations of FPOP analysis to characterize and localize regions of protein-protein interactions involving IDPs were illustrated.

In the Chapter 4, we discussed problems associated with FPOP techniques to study proteins. We also tested whether the excess concentration of a protein partner has an effect on the dosage of hydroxyl radical delivered to the target protein. On the basis of the data presented, we concluded that the extra protein sometimes interferes with the hydroxyl radical labeling. One of the reasons might be that extra protein consumes the hydroxyl radicals. In such cases, the reproducibility of the experiment will be affected because the reproducibility of FPOP experiments highly depends on the ability to deliver a defined concentration of hydroxyl radicals to the protein of interest. As a continuation of this research, two approaches can be explored to solve the excess concentration problem. In the first approach, a protein, which does not interact with the proteins under study, can be added to the protein mixture to compensate for the excess

protein concentration in the FPOP experiment and have the total protein concentration constant. For example, 4-fold molar excess of carbonic anhydrase, as a concentration normalizing additive, can be added to (1 ACTR: 4 CBP) complex, and 4-fold molar excess of carbonic anhydrase can be added to (4 ACTR: 1 CBP) complex to compensate for the excess concentration problem. Another approach to measure a defined concentration of hydroxyl radicals generated and delivered to the protein target is by introducing a radical dosimeter into the FPOP experiment. Adenine, which has been previously used as a radical dosimeter,[(20)] can be tested for ACTR and CBP experiment to normalize the FPOP data according to different oxidation conditions.

Another approach to improve the FPOP experiment to be able to characterize and determine the degree of structure in IDPs in their free state is by introducing a reference state. The idea of introducing a reference state is similar to the idea explained in Chapter 3 for millisecond hydrogen exchange. The only difference here for FPOP experiment, the amino acid side chains of the reference state would be unprotected from oxidation. For this experiment, since ACTR and CBP are small proteins, ^{15}N-labeled tryptic peptides of ACTR and CBP might be used as unprotected reference state to measure protection from oxidation of ACTR and CBP separately. This experiment would also need a radical dosimetry to reliably measure the oxidation, and adenine can be tested here as well.

5.3 References

1. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nature reviews. Molecular cell biology 6*, 197-208.

2. Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Obradovic, Z., and Uversky, V. N. (2007) Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins, *Journal of proteome research 6*, 1917-1932.

3. Tantos, A., Han, K.-H., and Tompa, P. (2012) Intrinsic disorder in cell signaling and gene transcription, *Molecular and cellular endocrinology 348*, 457-465.

4. Wright, P. E., and Dyson, H. J. (2009) Linking folding and binding, *Current opinion in structural biology 19*, 31-38.

5. Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J., Evans, R. M., and Wright, P. E. (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators, *Nature 415*, 549-553.

6. Lacy, E. R., Filippov, I., Lewis, W. S., Otieno, S., Xiao, L., Weiss, S., Hengst, L., and Kriwacki, R. W. (2004) p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding, *Nature structural & molecular biology 11*, 358-364.

7. Uversky, V. N. (2011) Intrinsically disordered proteins from A to Z, *The International journal of biochemistry & cell biology 43*, 1090-1103.

8. Receveur-Brechot, V., Bourhis, J. M., Uversky, V. N., Canard, B., and Longhi, S. (2006) Assessing protein disorder and induced folding, *Proteins 62*, 24-45.

9. Fitzgerald, M. C., and West, G. M. (2009) Painting proteins with covalent labels: what's in the picture?, *Journal of the American society for mass spectrometry 20*, 1193-1206.

10. Wales, T. E., and Engen, J. R. (2006) Hydrogen exchange mass spectrometry for the analysis of protein dynamics, *Mass spectrometry reviews 25*, 158-170.

11. Xu, G., and Chance, M. R. (2007) Hydroxyl radical-mediated modification of proteins as probes for structural proteomics, *Chemical reviews 107*, 3514-3543.

12. Rumi-Masante, J., Rusinga, F. I., Lester, T. E., Dunlap, T. B., Williams, T. D., Dunker, A. K., Weis, D. D., and Creamer, T. P. (2012) Structural basis for activation of calcineurin by calmodulin, *Journal of molecular biology 415*, 307-317.

13. Aramburu, J., Rao, A., and Klee, C. B. (2000) Calcineurin: from structure to function, *Current topics in cellular regulation 36*, 237-295.

14. Yang, S.-A., and Klee, C. B. (2000) Low Affinity Ca2+-Binding Sites of Calcineurin B Mediate Conformational Changes in Calcineurin A, *Biochemistry 39*, 16147-16154.

15. Zhang, H., Gau, B. C., Jones, L. M., Vidavsky, I., and Gross, M. L. (2011) Fast photochemical oxidation of proteins for comparing structures of protein-ligand complexes: the calmodulin-peptide model system, *Analytical chemistry 83*, 311-318.

16. Keppel, T. R., and Weis, D. D. (2013) Analysis of disordered proteins using a simple apparatus for millisecond quench-flow H/D exchange, *Analytical chemistry 85*, 5161-5168.

17. Kjaergaard, M., Norholm, A. B., Hendus-Altenburger, R., Pedersen, S. F., Poulsen, F. M., and Kragelund, B. B. (2010) Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II?, *Protein science : a publication of the protein society 19*, 1555-1564.

18. Ebert, M. O., Bae, S. H., Dyson, H. J., and Wright, P. E. (2008) NMR relaxation study of the complex formed between CBP and the activation domain of the nuclear hormone receptor coactivator ACTR, *Biochemistry 47*, 1299-1308.

19. Keppel, T. R., and Weis, D. D. (2015) Mapping residual structure in intrinsically disordered proteins at residue resolution using millisecond hydrogen/deuterium exchange and residue averaging, *Journal of the American society for mass spectrometry 26*, 547-554.

20. Xie, B., and Sharp, J. S. (2015) Hydroxyl Radical Dosimetry for High Flux Hydroxyl Radical Protein Footprinting Applications Using a Simple Optical Detection Method, *Analytical chemistry 87*, 10719-10723.